Eötvös Loránd University
Faculty of Informatics

Distribution of additive arithmetical functions

Ph.D. Dissertation

by

László Germán

Supervisor
Prof. Dr. Imre Kátai
member of the Hungarian Academy of Sciences

Doctoral School of Informatics
Director: Prof. Dr. János Demetrovics
member of the Hungarian Academy of Sciences

Doctoral Program: Numerical and Symbolical Computation
Program Director: Prof. Dr. Antal Járai
Doctor of Mathematical Sciences

2008.

Contents

Abstract

This thesis is devoted to the study of the distribution of additive arithmetical functions in some subset of natural numbers. This topic is known in mathematics as Probabilistic Number Theory.
The aim of this work is to generalize the results of Kátai [28] and Hildebrand [21] about the distribution of additive functions in the set of "shifted primes". It shows how to extend their work for much more general sequences, admitting some necessary properties. The method will be clarified through the example of the natural numbers for which we take the "shifts" of natural numbers composed from the product of a given number of primes. The involved probabilistic approach, and mean value estimates in the theory of arithmetical functions on this type of problems are generalized. An application is shown to Erdős-Kac type theorems. The theory of the involved methods are far from to be closed. For instance the mean behavior of multiplicative functions on other sequences than the whole natural numbers is not know. Although to answer these questions completely for the most simplest cases like the sequence of the "shifted primes" seams to be out of reach - due to the problem in analytic number theory known as the *parity obstacle* see [27] and [39] - one is able to get some results conditionally.

Acknowledgments

Above all I would like to thank my family for covering required circumstances to work. Especially to my parents for their extraordinary sacrifice, to my wife for her love, and to my sister for lots of help. It is to them that I dedicate this dissertation.
I would like to express my gratitude to Professor Imre Kátai for his honestly support and for his attention, to Professor Karl-Heinz Indlekofer for his support and for his usefull advices. A special thanks to other colleagues, friends for their assistance.

Chapter 1

Introduction

It is not easy to determine the roots of Probabilistic Number Theory. One could conjecture some probabilistic concepts, for example behind the Prime Number Theorem, which was conjectured by Legendre and the young Gauss around 1800. The theorem was established independently by Hadamard and de la Vallée Poussin in 1896. Although it is difficult to see the probabilistic aspects through the proofs, the Cramer model tells us, that we can think about the prime property being a random event. This model is rather heuristic, it is not known if there is a pure probabilistic approach concerning the distribution of prime numbers, which would give us the Prime Number Theorem.

The situation is slightly favorable if we consider additive arithmetical functions.

An arithmetical function $f : \mathbb{N} \to \mathbb{C}$ is said to be additive if $f(n \cdot m) = f(n) + f(m)$ whenever $(m, n) = 1$. Such a function is completely determined by its values on prime powers. If $f(n{\cdot}m) = f(n){+}f(m)$ holds unconditionally, then f is completely additive, and if $f(p^\alpha) = f(p)$ is constant for $\alpha \geq 1$, whenever p is a prime number, then f is said to be strongly additive. Real valued strongly additive function plays a special role in number theory. For these functions it is possible to give a pure probabilistic description. This is roughly speaking to approximate the distribution of additive arithmetical functions by the sum of independent random variables. We introduce two measures in the probability spaces on the sequence of natural numbers, one of which is the normalized counting function, and the other is the asymptotic density on a proper collection of subsets of $\mathbb{N}$. On the second probability space the distribution of sum of appropriate independent random variables describes the distribution of additive functions very well. After this we can apply the results of probability theory about the sum of independent random variables. The discussion of this approach is given in Chapter 3.

One relevant step in this direction was done by Erdős and Wintner (See [8] for example) by characterizing the set of additive arithmetical functions which obey a limit law on the set of natural numbers. Their result was as follows:
Let f be a real additive function, and $1 < x$. The frequency of f is defined by

$$F_x(z) := \frac{1}{x} \sum_{\substack{n \le x \\ f(n) \le z}} 1,$$

for all real z. Then $F_x(z)$ is a distribution function and we say that f has a limit distribution, if

$$F_x(z) \to F(z) \quad (x \to \infty),$$

in all z continuity points of F, where F is a distribution function as well. This type of convergence has been called weak convergence and we will denote it by

$$F_x \Rightarrow F \quad (x \to \infty).$$

It turned out, that the above property of f is equivalent to the following condition, which we call the Erdős-Wintner condition, i.e. the three series over prime numbers

$$\sum_{|f(p)|>1} \frac{1}{p}, \quad \sum_{|f(p)|\le 1} \frac{f(p)}{p}, \quad \sum_{|f(p)|\le 1} \frac{f^2(p)}{p}$$

converge.

The full probabilistic description with many applications was made by Kubilius in his monograph [26]. About this time Kátai [28] began to characterize the additive functions having a limit distribution on the set of "shifted primes".The full description of such functions was given by Hildebrand [21] about 20 years after. In susequent papers the method of Kátai was extended to more general subsets of integers for a wide class of additive functions ([29], [22]).

This problem can be investigated in a context more general than that of Erdős and Wintner. For this purposes let f be an additive function and A_x be a subset of $\mathbb{N}$ such that $A_x \cap [1..x]$ is nonempty, with $1 < x$. Then the frequency of f on A_x is now defined by

$$\nu_x(n \in A_x;\ f(n) \le z) := \frac{1}{|A_x \cap [1..x]|} \sum_{\substack{n \le x \\ n \in A_x \\ f(n) \le z}} 1.$$

Then one can ask if the frequencies possess a limit law or not, that is if

$$\nu_x(n \in A_x;\ f(n) \le z) \Rightarrow F(z) \quad (x \to \infty) \tag{1.1}$$

or not with a distribution function $F(z)$. Erdős and Wintner considered this question taking A_x to be $\mathbb{N}$. Kátai and Hildebrand investigated the case $A_x = \mathcal{P} + 1$, where $\mathcal{P}$ denotes the set of primes.

As a typical application of the above results like in [8] we conclude for the sum of divisors function σ that

$$\#\{n \le x:\ \sigma(n) \le zn\} = h(z)x + o(1) \quad (x \to \infty),$$

where $h(z)$ is a distribution function. The main topic of the thesis is to investigate this type of distribution problems if we take

$$A_x = \{n \le x:\ \omega(n-1) = k_x\},$$

where $\omega(n)$ denotes the number of distinct prime factors of n, and $1 \le k_x \le \varepsilon(x)\sqrt{\log\log x}$ with $\varepsilon(x) \to 0 \quad (x \to \infty)$. Note that the problem of Kátai and Hildebrand corresponds to the case $k_x = 1$ (apart from the higher prime powers, which have zero relative density in the set of prime powers). The detailed discussion of this problem after technical preparation in Chapter 2 will be given in Chapter 4.

One way to handle the problem described above, is to give a pure probabilistic description of arithmetical functions as in Chapter 3. Alternatively one can solve the problem via characteristic functions. This method is related to mean value estimates of multiplicative functions and we will discuss it in Chapter 5.

A further application of the results discussed in Chapter 3 allows us to give an Erdős-Kac type Theorem in Chapter 6 which seems to be stated in the literature only for fixed k, and with further strong restrictions on the set A_x (See [2]).

Chapter 2

Technical preparation

To state the problem and to discuss the results we need to do some technical consideration. We will use some basic facts in number theory, about the distribution of prime numbers in arithmetical progressions. We will also use results in probability theory.

Throughout this work n denotes a positive integer and $P(n)$, $p(n)$ denotes the largest and the smallest prime factor of n respectively. p, q with or without suffixes will always denote prime numbers. As usually the number of primes up to x will be denoted by $\pi(x)$, and $\log x$ means the natural logarithm of x. For the number of primes up to x in arithmetic progression we will use the notation $\pi(x, d, l)$ that is, if $(d, l) = 1$ then

$$\pi(x, d, l) = \sum_{\substack{p \leq x \\ p \equiv l \pmod d}} 1.$$

If

$$n = p_1^{r_1} \cdot p_2^{r_2} \cdots p_k^{r_k}, \qquad p_1 < p_2 < \ldots < p_k \tag{2.1}$$

where p_i, $i = 1, .., k$ are distinct primes then let $\omega(n) := k$. A typical integer n for which $\omega(n) = k$ will be denoted by π_k. We denote the set of integers having k distinct prime factor with $\mathcal{P}_k$ that is

$$\mathcal{P}_k := \{\pi_k \in \mathbb{N}\}.$$

The set of integers in $\mathcal{P}_k$ up to x is denoted by $\mathcal{P}_k(x)$. We introduce the counting function for the set $\mathcal{P}_k$ in arithmetic progressions. If $(d, l) = 1$ then let

$$\pi_k(x, d, l) = \sum_{\substack{\pi_k \leq x \\ \pi_k \equiv l \pmod d}} 1.$$

In the special case $d = l = 1$ we use $\pi_k(x)$ instead of $\pi_k(x, 1, 1)$ and if $k = 1$ then we use $\pi(x, d, l)$ instead of $\pi_1(x, d, l)$. In what follows we will use some facts about the structure of $\mathcal{P}_k$ as well as about the distribution of integers in $\mathcal{P}_k$ in arithmetic progressions. First we introduce the results in the well studied case, where $k = 1$ and then we seek to the general case, where k can vary in a wide range depending on x.

The following result can be found for example in [34] (Satz V.2.1).

Lemma 1 (Brun-Titchmarsh)**.** *Let x be a real number, and let $0 < a < 1$. Let k be positive integer with $1 \le k \le x^a$. Then we have*

$$\pi(x, k, l) < c(a)\frac{x}{\varphi(k)\log x},$$

for all $1 \le l < k$, $(k, l) = 1$.

For a general modulus we can use (Satz II.4.1 in [34])

Lemma 2. *Let $1 \le k < x$, $0 \le l < k$, $(k, l) = 1$. Then*

$$\pi(x, k, l) < c\frac{x}{\varphi(k)\log\left(\frac{x}{k}\right)},$$

where c is a uniform constant.

If the modulus is small in terms of x then we are able to give the precise value of $\pi(x, k, l)$ with a small error term. To describe the main term we need the following notation. Let

$$li(x) := \int_2^x \frac{1}{\log u}du.$$

With this notation we have the following result. See [34] (Satz IV.8.3).

Lemma 3 (Siegel-Walfisz)**.** *Let A be a positive real number. We have*

$$\pi(x, k, l) = \frac{1}{\varphi(k)}li(x) + \mathcal{O}\left(xe^{-c\sqrt{\log x}}\right)$$

uniformly for all $1 \le k \le \log^A x$, $(k, l) = 1$, $1 \le l < k$.

Although the result above fails, if the modulus is large in terms of x, we are able to give the value of $\pi(x, k, l)$ with small error term in average for large modulus. An elegant proof of the result can be found in [6] (Chapter 28.)

Lemma 4 (Bombieri-Vinogradov). *Let $A > 0$ be fixed. Then*

$$\sum_{d \leq x^{1/2} \log^{-A} x} \max_{y \leq x} \max_{(a,d)=1} \left| \psi(y,d,a) - \frac{y}{\varphi(d)} \right| \ll x \log^{-A+5} x,$$

where

$$\psi(x,d,a) = \sum_{\substack{p^\alpha \leq x \\ p^\alpha \equiv a \pmod{d}}} \log p.$$

Remark. *Since clearly*

$$\psi(x,d,a) \ll \frac{x}{d} \log x$$

for $d \leq x$, therefore using a simple partial summation one can easily conclude that for all $A > 0$ there exists a $B > 0$ such that

$$\sum_{d \leq x^{1/2} \log^{-B} x} \max_{y \leq x} \max_{(a,d)=1} \left| \pi(y,d,a) - \frac{li(y)}{\varphi(d)} \right| \ll x \log^{-A} x.$$

The corresponding result concerning multiplicative functions was given in [49].

We will use the following well known asymptotic about the sum of reciprocals over primes in arithmetic progressions, which follows easily from Lemma 3. by a partial summation.

Lemma 5 (Dirichlet). *With $A > 0$ we have*

$$\sum_{p \equiv l \pmod{k}} \frac{1}{p} = \frac{1}{\varphi(k)} \log\log x + C_k + \mathcal{O}\left(\frac{1}{\log x}\right)$$

uniformly for all $1 \leq k \leq \log^A x$, $(k,l) = 1$, $1 \leq l < k$.

The following result is a direct consequence of the above Lemma. (See for example [32].)

Lemma 6 (Mertens). *For $x \geq 2$*

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right) = B \log x + \mathcal{O}(1),$$

where $B > 0$.

The following well known result can be found for example in [40], [36], [37].

Lemma 7 (Sathe-Selberg)**.** *Let $A > 0$, $x \geq 3$. Then we have with*

$$\lambda(z) := \frac{1}{\Gamma(z+1)} \prod_p \left(1 + \frac{z}{p-1}\right)\left(1 - \frac{1}{p}\right)^z$$

that

$$\pi_k(x) = \frac{x}{\log x} \frac{(\log\log x)^{k-1}}{(k-1)!} \left\{ \lambda\left(\frac{k-1}{\log\log x}\right) + \mathcal{O}\left(\frac{k}{(\log\log x)^2}\right)\right\}$$

uniformly for all $1 \leq k \leq A \log\log x$ where the implied constant depends only on A.

We let k vary in a wide range in the dependency of x. To describe this domain for k we introduce the following definition: Let $\varepsilon(x)$ be a function such that $\varepsilon(x) \to 0$ as $x \to \infty$. We say that k has property $A(\varepsilon, x)$ if $2 \leq k \leq \varepsilon(x)\sqrt{\log\log x}$.

Suppose that n is in the form (2.1). Let

$$\delta_j(n) = p_1^{r_1} \cdots p_j^{r_j} \quad (j = 1, \ldots, k),$$

and

$$\gamma_j(n) = \frac{\log p_{j+1}}{\log \delta_j(n)} \quad (j = 1, \ldots, k-1).$$

The first result we prove, shows that for almost all integers in $\mathcal{P}_k$ the size of the $j+1$-th prime factor is large in terms of the product of the first j prime factor of n. That is, if

$$\Delta(n) = \min_{j=1,\ldots k-1} \gamma_j(n)$$

than we have the following

Lemma 8. *Suppose k has property $A(\varepsilon, x)$ and let $M = M_x$ be defined by $\log\log M = \sqrt{\log\log x}$. Then there exists a sequence A_x, depending at most on $\varepsilon(x)$ with $A_x \to \infty$ as $x \to \infty$ such that*

$$\Delta(\pi_k) \geq A_x,$$

and

$$r_1 = r_2 = \cdots = r_k = 1, \quad p_1 > M$$

holds for all but $o(1)\pi_k(x)$ element of $\mathcal{P}_k(x)$ as $x \to \infty$.

Proof. First we estimate the number of integers in $\mathcal{P}_k(x)$, which are not squarefree. We have with

$$A_1 := \{n \in \mathcal{P}_k(x) : \exists p^2 | n\}, \tag{2.2}$$

that

$$\#A_1 \leq \sum_{p^2 \leq x^{1/2}} \pi_{k-1}\left(\frac{x}{p^2}\right) + \sum_{p^2 > x^{1/2}} \frac{x}{p^2}. \tag{2.3}$$

The second term here is $\mathcal{O}(x^{3/4})$. To estimate the first term we note that, since k has property $A(\varepsilon, x)$, Lemma 7. applies, and we get

$$\frac{\pi_{k-1}(x)}{\pi_k(x)} \sim \frac{k}{\log\log x} \quad (x \to \infty),$$

such that the first therm on the right hand side of (2.3) is at most

$$c\pi_k(x)\frac{k}{\log\log x}\sum_{p^2 \leq x^{1/2}} \frac{1}{p^2} \ll o(\pi_k(x)).$$

We now seek an estimate to the number of integers in $\mathcal{P}_k(x)$, for which there is a small prime factor. Let

$$A_2 := \{n \in \mathcal{P}_k(x) : p(n) \leq M\}.$$

Then we have that

$$\#A_2 \leq \#A_1 + \sum_{\substack{p \leq x^{1/2} \\ p < M}} \pi_{k-1}\left(\frac{x}{p}\right)$$

We have seen that the first term here is $o(\pi_k(x))$. The second term does not exceed

$$c\pi_k(x)\frac{k}{\log\log x}\sum_{p \leq M}\frac{1}{p},$$

which by the definition of M is at most

$$\pi_k(x)\frac{k}{\sqrt{\log\log x}},$$

therefore it is $o(\pi_k(x)$. From now on we can assume that for all but $o(\pi_k(x))$ element n in $\mathcal{P}_k(x)$ we have that n is squarefree and that $p(n) > M$. Now we consider the integers n in $\mathcal{P}_k(x)$ for which

$$\Delta(n) > A_x.$$

For such an n we have that if $pq|n$, $p < q$ then

$$p^{A_x} \leq q.$$

The number of integers n in $\mathcal{P}_k(x)$ for which $n \notin A_1$ and $n \notin A_2$ and $q < p^{A_x}$ whenever $pq|n$, $p < q$ is not more than

$$\sum_{\substack{pq\pi_{k-2} \leq x \\ M<p<q<p^{A_x} \\ p \leq x^{\frac{1}{2A_x}}}} 1 + \sum_{\substack{pq\pi_{k-2} \leq x \\ M<p<q<p^{A_x} \\ p > x^{\frac{1}{2A_x}}}} 1. \tag{2.4}$$

The first sum on the right hand side is as above at most

$$\sum_{\substack{M<p<q<p^{A_x} \\ p \leq x^{\frac{1}{2A_x}}}} \pi_{k-2}\left(\frac{x}{pq}\right) \ll \frac{x}{\log x} \frac{(\log\log x)^{k-3}}{(k-3)!} \sum_{\substack{M<p<q<p^{A_x} \\ p \leq x^{\frac{1}{2A_x}}}} \frac{1}{pq}$$

$$\ll \pi_k(x) \frac{k^2}{\log\log x} \log A_x,$$

which is $o(\pi_k(x))$ if A_x tends to infinity suitable slowly as $x \to \infty$. To estimate the second term in (2.4) we note that the number of integers up to x for which $p(n) > z$ is at most

$$\frac{x}{\log z} + c\frac{z^2}{\log^2 z}$$

if $z \geq z_0$. This result can be found for example in [16] (Theorem 3.6). The sum we want to estimate is therefore

$$\sum_{\substack{pq\pi_{k-2} \leq x \\ M<p<q<p^{A_x} \\ p > x^{\frac{1}{2A_x}}}} 1 \ll \sum_{\pi_{k-2} \leq x^{1-\frac{1}{A_x}}} \sum_{\substack{n \leq \frac{x}{\pi_{k-2}} \\ p(n) > x^{\frac{1}{2A_x}}}} 1$$

$$\ll A_x \frac{x}{\log x} \sum_{\pi_{k-2} \leq x^{1-\frac{1}{A_x}}} \frac{1}{\pi_{k-2}} + A_x^2 \frac{x^{\frac{1}{A_x}}}{\log x} \sum_{\pi_{k-2} \leq x^{1-\frac{1}{A_x}}} 1 \tag{2.5}$$

Since

$$\sum_{\pi_{k-2}\le x}\frac{1}{\pi_{k-2}} \ll \frac{1}{(k-2)!}\left(\sum_{p^\alpha\le x}\frac{1}{p^\alpha}\right)^{k-2}$$
$$\ll c\left(\sum_{p\le x}\frac{1}{p}+c\right)^{k-2}$$
$$\ll \frac{(\log\log x)^{k-2}}{(k-2)!}$$

if x is large enough, therefore the right hand side of (2.5) is at most $o(\pi_k(x))$ providing A_x tends to infinity appropriately slowly. □

Remark. *We are considering distribution problems in the meaning of* (1.1)*, therefore we can identify $\pi_k(x)$ with the set containing integers with the property described in the above lemma.*

In the sense of this remark we introduce the following notations. Let

$$U_k(x) := \{n \in \mathcal{P}_k(x) : \Delta(n) \ge A_x,\ \mu^2(n) = 1, \ \text{ and } \ p(n) > M\},$$

and

$$B_k(x) := \#U_k(x).$$

We need some knowledge about the distribution of integers from $\mathcal{P}_k$ in arithmetic progression. For this purpose we will use some analogous from the theory of prime numbers. With the notation

$$B_k(x,d,l) := \sum_{\substack{n\in U_k(x)\\ n\equiv l \pmod q}} 1$$

whenever $(d,l)=1$, we have the following

Lemma 9. *Let $x \ge x_0$ and let $d \le \log^A x$, where A is a positive fixed number. Let furthermore $(d,l)=1$, and let k be a positive integer having property $A(\varepsilon,x)$. Then we have uniformly for all k with that property that*

$$B_k(x,d,l) = \frac{B_k(x)}{\varphi(d)}\left(1+\mathcal{O}\left(e^{-c\sqrt{\log x}}\right)\right).$$

Proof. Let S_x be the set of those π_{k-1}, for which there exists at least one prime $p > P(\pi_{k-1})$ such that $\pi_{k-1}p \in U_k(x)$. Let $p^* = p_{\pi_{k-1}}$ be the smallest p with this property. Then $\pi_{k-1}p \in U_k(x)$ for all $p^* \le p \le \frac{x}{\pi_{k-1}}$. Let

$$\pi_{k-1}p \equiv l \pmod{d}.$$

Then, using Lemma 8., $\pi_{k-1} < x^\lambda$, with an appropriate $\lambda < 1/4$,

$$P\left(\pi_{k-1}\right) < p, \quad \text{and} \quad p\left(\pi_{k-1}\right) > \log^A x,$$

when x is larger than x_A say. With this conditions $(\pi_{k-1}, d) = 1$, therefore

$$p \equiv l_{\pi_{k-1}} \pmod{d}$$

with a unique $l_{\pi_{k-1}} \mod d$.
We have that

$$\left| B_k\left(x, d, l\right) - \frac{B_k\left(x\right)}{\varphi\left(d\right)} \right| \ll \sum_{\pi_{k-1} \in S_x} \left| E_1 - E_2 \right|, \tag{2.6}$$

where

$$E_1 = \pi\left(\frac{x}{\pi_{k-1}}, d, l_{\pi_{k-1}}\right) - \frac{1}{\varphi\left(d\right)}\pi\left(\frac{x}{\pi_{k-1}}\right),$$

and

$$E_2 = \pi\left(p^*\left(\pi_{k-1}\right), d, l_{\pi_{k-1}}\right) - \frac{1}{\varphi\left(d\right)}\pi\left(p^*\left(\pi_{k-1}\right)\right).$$

Noting that

$$M_x \le p^*(\pi_{k-1}) \le \frac{x}{\pi_{k-1}},$$

the Theorem of Siegel-Walfisz for prime numbers is applicable to

$$\sum_{\pi_{k-1} \in S_x} \left| \pi\left(\frac{x}{\pi_{k-1}}, d, l_{\pi_{k-1}}\right) - \frac{1}{\varphi\left(d\right)}\pi\left(\frac{x}{\pi_{k-1}}\right) \right|,$$

and we deduce that the right hand side of (2.6) does not exceed

$$\frac{c}{\varphi\left(d\right)} \sum_{\pi_{k-1} \in S_x} \frac{x}{\pi_{k-1}} e^{-c\sqrt{\log x}} + x^\lambda \sum_{\pi_{k-1} \in S_x} 1.$$

The sum appearing in the first term of this last expression can be estimated as follows:

$$\sum_{\pi_{k-1}\in S_x} \frac{1}{\pi_{k-1}} \le \frac{1}{(k-1)!}\left(\sum_{p\le x}\frac{1}{p}\right)^{k-1},$$

since after extracting the product on the right hand side, all the terms

$$\frac{1}{p_1\cdot p_2\cdots p_{k-1}}$$

with different p_i-s, appear exactly $(k-1)!$ times. Furthermore

$$\left(\sum_{p\le x}\frac{1}{p}\right)^{k-1} \ll \log\log^{k-1} x,$$

which follows since k has property $A(\varepsilon,x)$. □

Remark. *We note that in the sense of Lemma 8. we have*

$$\pi_{k-1}^{A_x} \le p^* \le (1+\varepsilon)\pi_{k-1}^{A_x} \le x$$

for all fixed ε and large enough x. But this lemma remains true if we take $3A_x$ instead of A_x, therefore $U_k(x)$ can be chosen such that $p^* \le x^{1/2}$ hold for all $\pi_{k-1}\in S_x$.

Let

$$B_k\left(x|d\right) = \sum_{\substack{n\in U_k(x)\\ (n,d)=1}} 1\,.$$

Lemma 10. *Let $x\ge 2$, $A>0$ an arbitrary number, and $\alpha<1/2$. We have with all k having property $A(\varepsilon,x)$ that*

$$\sum_{d<x^\alpha} \max_{(l,d)=1}\max_{z\le x}\left|B_k\left(z,d,l\right) - \frac{B_k\left(z|d\right)}{\varphi\left(d\right)}\right| \ll B_k\left(x\right)\log^{-A} x.$$

The constant implied by $\ll$ does not depend on k.

Proof. With the notations of Lemma 9. we have for $x>x_0$ that

$$\left|B_k\left(x,d,l\right) - \frac{B_k\left(x|d\right)}{\varphi\left(d\right)}\right| \le \sum_{\substack{\pi_{k-1}\in S_x\\ (\pi_{k-1},d)=1}} \left|\pi\left(\frac{x}{\pi_{k-1}},d,l_{\pi_{k-1}}\right) - \frac{\pi\left(\frac{x}{\pi_{k-1}}|d\right)}{\varphi\left(d\right)}\right| + \mathcal{O}\left(x^{1/3}\right).$$

Applying the Bombieri-Vinogradov theorem we get that the right hand side of this last inequality, does not exceed

$$c\frac{x}{\log x}\log^{-A}x\sum_{\pi_{k-1}\in S}\frac{1}{\pi_{k-1}}+\mathcal{O}\left(x^{2/3}\right)\ll B_k\left(x\right)\log^{-A}x.$$

Here we used, that $\pi(x)-\pi(x|d)\leq\omega(d)\leq\log x$. $\square$

The asymptotic concerning the distribution of integers in $\mathcal{P}_k$ in progressions was investigated in a slightly different context in [42], [43]. A more general version of lemma 10 was established by Wolke and Zhang [50].

Lemma 11 (Wolke-Zhang)**.** *For any given $A>0$ and $0<\varepsilon<1/2$ there exist $\eta>0$ such that*

$$\sum_{d\leq x^{1/2-\varepsilon}}\max_{(a,d)=1}\max_{y\leq x}\left|\pi_k\left(y,d,a\right)-\frac{\pi_k\left(y|d\right)}{\varphi\left(d\right)}\right|\ll\pi_k\left(x\right)\log^{-A}x$$

holds uniformly for $k\leq\eta\log x/\log\log^2 x$, and the constant implied in $\ll$ depends on A and ε only.

Lemma 12. *Let $0\leq\eta<1/4$. We have*

$$\sum_{x^\eta<q<x^{2\eta}}B_k\left(x,q,-1\right)=\sum_{x^\eta<q<x^{2\eta}}\frac{B_k\left(x\right)}{q-1}+\mathcal{O}\left(\frac{B_k\left(x\right)}{\log^A x}\right).$$

Proof. From Lemma 10. we have

$$\sum_{x^\eta<q<x^{2\eta}}B_k\left(x,q,-1\right)=\sum_{x^\eta<q<x^{2\eta}}\frac{B_k\left(x|q\right)}{q-1}+\mathcal{O}\left(\frac{B_k\left(x\right)}{\log^A x}\right).$$

On the other hand

$$B_k(x)\geq B_k\left(x|q\right)\geq B_k\left(x\right)-\sum_{l=1}^{\infty}\sum_{\substack{aq^l<x\\ a\in P_{k-1}(x)\\ (a,q)=1}}1\geq B_k\left(x\right)+\mathcal{O}\left(\frac{x}{q}\right).$$

Putting it together

$$\begin{aligned}\sum_{x^\eta<q<x^{2\eta}}\frac{B_k\left(x|q\right)}{q-1}&=\sum_{x^\eta<q<x^{2\eta}}\frac{B_k\left(x\right)}{q-1}+\mathcal{O}\left(\sum_{x^\eta<q<x^{2\eta}}\frac{x}{q^2}\right)\\&=\sum_{x^\eta<q<x^{2\eta}}\frac{B_k\left(x\right)}{q-1}+\mathcal{O}\left(x^{1-\eta}\right).\end{aligned}$$

$\square$

We will use the Fundamental Lemma of Sieve Theory in the following form ([8] Lemma 2.1.)

Lemma 13. *Let $a_n, n = 1, \dots, N$ be sequence of natural numbers. Let r be a positive real number, and let $p_1 < p_2 < \cdots < p_s \le r$ be prime numbers. Set $Q = p_1 p_2 \cdots p_s$. If $d|Q$ then let*

$$\sum_{\substack{n=1 \\ a_n \equiv 0 \pmod d}}^{N} 1 = \eta(d)X + R(N, d),$$

where X, $R(N, d)$ are real numbers with $X \ge 0$ and η is a multiplicative function. Suppose that

$$0 \le \eta(p) < 1$$

for all prime p. Then the estimate

$$\sum_{\substack{n=1 \\ (a_n, Q)=1}}^{N} 1 = \{1 + 2\theta_1 H\} X \prod_{p|Q} (1 - \eta(p)) + 2\theta_2 \sum_{\substack{d|Q \\ d \le z^3}} 3^{\omega(d)} |R(N, d)|$$

holds uniformly for $r \ge 2$, $\max(\log r, S) \le 1/8 \log z$, where $|\theta_1| \le 1, |\theta_2| \le 1$, and

$$H = \exp\left(-\frac{\log z}{\log r} \left\{ \log\left(\frac{\log z}{S}\right) - \log\log\left(\frac{\log z}{S}\right) - \frac{2S}{\log z} \right\} \right),$$

$$S = \sum_{p|Q} \frac{\eta(p)}{1 - \eta(p)} \log p$$

An interesting corollary of the above result is the following

Lemma 14.

$$\sum_{\sqrt{x} \le p \le x} p\pi^2(x, -1, p) \ll \pi^2(x).$$

For a proof see [8] Lemma 4.19. An other important result from Sieve Theory is an estimation about the number of primes in multiple arithmetic progressions (Satz II.4.2 in [34]).

Lemma 15. *Let s be a positive integer and let*

$$(a_i, b_i), \quad i = 1, \dots s$$

pairs of integers with

$$(a_i, b_i) = 1, \quad i = 1, \dots, s.$$

Let $\eta(p)$ denote the number of the $\pmod p$ solutions of the congruence

$$\prod_{i=1}^{s}(a_i m + b_i) \equiv 0 \pmod p,$$

and assume that

$$\eta(p) < p \quad \textit{for all primes } p.$$

Let furthermore

$$E = \prod_{i=1}^{s} a_i \prod_{1 \le r < k \le s} (a_r b_k - a_k b_r) \neq 0.$$

Then for $x \ge 2$,

$$\#\{n \le x, |a_i n + b_i| \textit{ prime for } i = 1, \dots s\} \ll \frac{x}{\log^s x} \prod_{p|E} \left(1 - \frac{1}{p}\right)^{-(s-\eta(p))},$$

where the implied constant depends on s only.

Chapter 3

Probabilistic approach

The probabilistic concept of additive arithmetical functions was described by Kubilius in [26] with many applications. The main idea is to consider the corresponding truncated function, for which the distribution can be approximated by the distribution of sum of independent random variables. The distribution of such sums is well studied in probability theory, such that it can be applied to the theory of additive functions. In this Chapter we discuss this approach shortly. Another way to investigate the distribution of additive functions is the characteristic function method described by Lévy , which is closely related to mean value estimations for multiplicative functions. Since such estimations are useful in many applications, we proceed by means of this last method in subsequent Chapters. The essence of the work of Lévy is the following result (See for example in [8]):

Lemma 16 (Lévy 's continuity theorem)**.** *Let $F_y(z)$ be a sequence of distribution functions, with characteristic functions $\phi_y(t)$. Then there is a distribution function F such that*

$$F_y(z) \to F(z) \quad (y \to \infty)$$

for all continuity points z of F, if and only if there is a function ϕ in the real variable t which is continuous at $t = 0$ such that

$$\phi_y(t) \to \phi(t) \quad (y \to \infty)$$

holds uniformly for all bounded values of t. In this case ϕ will be a characteristic function with the corresponding distribution function F.

Now if the set A_x in Chapter 1 is not empty, then there are distribution functions on the left hand side of (1.1) as well. The characteristic functions of these distributions are in the form

$$\frac{1}{|A_x \cap [1..x]|} \sum_{\substack{n \in A_x \\ n \leq x}} e^{itf(n)}, \tag{3.1}$$

therefore this last result allows us to transfer the question stated in (1.1) to the investigation of the limit behavior of (3.1).

In what follows, we set A_x to be $\mathcal{P}_k + 1$ with k having the property $A(\varepsilon, x)$ for some $\varepsilon(x)$. The theorem of Sathe and Selberg insures that $\mathcal{P}_k(x)$ is not empty if x is large enough (See for example [40]), such that

$$\nu_x(n \in A_x : f(n) \leq z) \tag{3.2}$$

is a distribution function for all x being large enough.

In the remainder of this Chapter we describe the probabilistic nature of additive arithmetical functions. This approach was first developed by Kac [24] and Kubilius [26], and subsequently by Galambos [14]. Here we follow the work of Billingsley [3], and Elliott [8], [9].

Let $m \in \mathbb{N}$ with

$$m = \prod p^{\beta_p(m)},$$

where $\beta_p(m)$ is the exact power of p in the canonical factorization of m. Then for all additive f and $m \in \mathbb{N}$ we have

$$f(m) = \sum_p f(p^{\beta_p(m)}).$$

If we set

$$\delta_p(m) = \begin{cases} 1 \text{ if } \beta_p(m) > 0 \\ 0 \text{ otherwise }, \end{cases}$$

then all strongly additive f are of the form

$$f(m) = \sum_p f(p)\delta_p(m).$$

Let $\nu_k(x)$ be the probability measure on the positive integers with point mass $\pi_k^{-1}(x)$ for each $m \in \mathcal{P}_k(x) + 1$. If we now set $\mathcal{A}_k(x)$ to be the set of all subsets of $\mathcal{P}_k(x) + 1$ then

$$S_k(x) = (\mathcal{P}_k(x) + 1, \mathcal{A}_k(x), \nu_k(x))$$

forms a finite probability space. Furthermore if $p_1 < p_2 \ldots < p_r$ are distinct primes then

$$\nu_k(x)(m : \delta_{p_i}(m) = 1,\ i = 1 \ldots r) = \pi_k(x)^{-1} \pi_k(x, p_1 p_2 \cdots p_r, -1),$$

which by Lemma 9 equals

$$\frac{1}{\varphi(p_1 p_2 \cdots p_r)} + o(1)$$

as x tends to infinity, i.e. the random variables δ_p are almost-independent to one another.

The distribution in (3.2) then equals to the distribution of the random variable

$$\xi = \sum_p f(p) \delta_p.$$

In the contrary to independent random variables, the almost-independent random variables do not have some of the convenient properties - like the linearity of the expectation of sums. However in Chapter 4 we can prove the following inequality

$$\sum_{n \in \mathcal{P}_k(x)} \left| f_r(n+1) - \sum_{p \le r} \frac{f(p)}{p} \right|^2 \ll \pi_k(x) \sum_{p \le r} \frac{f^2(p)}{\varphi(p)},$$

where $f_r(n)$ is the truncated version of $f(n)$. This type of inequalities has been called Turán-Kubilius inequality.

By truncating f we can approximate ξ with sum of independent random variables, which are of simpler nature, and there are several results which can be applied immediately for such sums. We define the independent random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ by their distribution function as follows. Let

$$X_p = \begin{cases} 1 \text{ with probability } \frac{1}{\varphi(p)} \\ 0 \text{ with probability } 1 - \frac{1}{\varphi(p)} \end{cases}$$

for all primes p. Note that the existence of the independent random variables with given distribution on some probability space, is a consequence of Kolmogorov's existence theorem (See [4] Theorem 20.4,[35]). Using the inclusion-exclusion principle and Lemma 9 we have by the independence that

$$\begin{aligned} \nu_k(x)(\delta_{p_i} = k_i,\ k_i \in \{0,1\},\ i = 1 \ldots r) =& \mathbb{P}(X_{p_i} = k_i,\ k_i \in \{0,1\},\ i = 1 \ldots r) \\ &+ \mathcal{O}(\exp(-c\sqrt{\log x})) \end{aligned}$$

uniformly for all $x \gg 1$ and $p_1, p_2, \dots, p_r \le \log\log x$.

Define the truncated function f_r by

$$f_r(p) = \begin{cases} f(p) \text{ if } p \le r \\ 0 \text{ otherwise} \end{cases},$$

for $r = \log\log x$.

Putting it all together we have that

$$\nu_k(x)(f_r(m) \le z) = \mathbb{P}\left(\sum_{p \le r} X_p \le z\right) + \mathcal{O}\left(\exp(-c\sqrt{\log x})\right).$$

Although there are more sophisticated probabilistic models concerning additive arithmetical functions, this particular model would be enough to get sufficient conditions for the existence of the limit law based only on probabilistic concepts. The level of truncation ($r = \log\log x$) needed here does not allow us to obtain general result, analogous to the central limit theorem for example. To avoid this difficulty however in Chapter 6 we will use the following

Lemma 17. *Let $f(n)$ ba a real strongly additive arithmetical function, $x > 2, 0 < \sigma \le 1$ and let*

$$f_r(n) = \sum_{\substack{p|n \\ p \le r}} f(p),$$

for some $r \le x^\sigma$. Let

$$K_D(x) = \{Dp + 1 \le x : p \in \mathcal{P}\}.$$

Then with an arbitrary constant $A > 0$ we have

$$(\nu_x =)\nu(n \in K_D(x) : f_r(n) \le z) = \mathbb{P}(\sum_{\substack{p \le r \\ p \nmid D}} X_p \le z) + \mathcal{O}(\exp(-\frac{\log x}{\log r}) + \log^{-A} x) \tag{3.3}$$

uniformly for all $1 \le D \le x^{1-\sigma}$, where for all primes p the X_p's are independent random variables over some appropriate probability space $(\Omega, \mathcal{A}, \mathbb{P})$ defined by their distribution as

$$X_p = \begin{cases} f(p) & \textit{with probability } \frac{1}{\varphi(p)} \\ 0 & \textit{with probability } 1 - \frac{1}{\varphi(p)}. \end{cases}$$

Proof. We build up two propability spaces. Let $(\mathcal{P}_D =)\mathcal{P}(r) = \prod_{\substack{p \le r \\ p \nmid D}} p$. As usual, for a prime p let

$$E(p) = \{n \in K_D(x) : n \equiv 0 \pmod p\},$$

and

$$\overline{E}(p) = K_D(x) \setminus E(p).$$

Furthermore for a divisor d of $\mathcal{P}(r)$ let

$$E_d = \bigcap_{p|d} E(p) \bigcap_{p|\mathcal{P}(r)/d} \overline{E}(p).$$

It is clear that the E_d's are distinct for different values of d, and on the left hand side of (3.3) it appears a measure of a set which is union of finitely many of them. Therefore if we take $\mathcal{A}$ to be the smallest σ-algebra which contains all the E_d's, then $f_r(n)$ is a random variable on the probability space $(K_D(x), \mathcal{A}, \nu)$. We compute the measure of a typical $A \in \mathcal{A}$. Let

$$\delta_d(A) = \begin{cases} 1 & \text{if } A \cap E_d \neq \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

We can then write

$$\nu A = \sum_d \delta_d(A) \nu(E_d). \tag{3.4}$$

For each divisor t of $\mathcal{P}(r)$ we define $(R_D =)R(x,t)$ by the identity

$$\sum_{\substack{n \in K_D(x) \\ n \equiv 0 \pmod t}} 1 = \frac{\pi(\frac{x-1}{D})}{\varphi(t)} + R(x,t).$$

Since for a divisor d of $\mathcal{P}(r)$ we have

$$\nu E_d = \pi^{-1}(\frac{x-1}{D}) \sum_{\substack{n \in K_D(x) \\ d|n \\ (n, \mathcal{P}(r)/d)=1}} 1,$$

therefore according to Lemma 13., with the notation $\eta(p) = \varphi^{-1}(p)$ we obtain for all $d|\mathcal{P}(r)$ with $(\mathcal{P}(r)/d, 2) = 1$ and all $z \ge 8 \max(\log r, S)$, with

$$\begin{aligned} S = &\sum_{2 \le p \le r} \frac{\eta(p)}{1-\eta(p)} \log p \\ \le &2 \sum_{2 \le p \le r} \frac{\log p}{p} + \mathcal{O}(1) \\ \le &2 \log r + \mathcal{O}(1), \end{aligned}$$

that

$$\nu E_d = (1+2\theta_1 H)\varphi^{-1}(d) \prod_{p|\mathcal{P}(r)/d} (1-\frac{1}{\varphi(p)})$$

$$+ 2\theta_2 \pi^{-1}(\frac{x-1}{D}) \sum_{\substack{t|\mathcal{P}(r)/d \\ t\le z^3}} 3^{\omega(t)} |R(x,td)|, \tag{3.5}$$

where

$$H = \mathcal{O}(\exp(-c\frac{\log z}{\log r})),$$

$|\theta_1|, |\theta_2| \le 1$. Note that the above formula for νE_d remains valid if $2|\mathcal{P}(r)/d$, since in this case $\nu E_d = 0$. The measure

$$\mathbb{P}E(p) := \varphi^{-1}(p)$$

with the property

$$\mathbb{P}(E(p) \cap E(q)) = \varphi^{-1}(p)\varphi^{-1}(q)$$

is well defined on $\mathcal{A}$, and it is easy to see that

$$\mathbb{P}E_d = \varphi^{-1}(d) \prod_{p|\mathcal{P}(r)/d} (1-\frac{1}{\varphi(p)}). \tag{3.6}$$

Since

$$\mathbb{P}K_D(x) = \sum_{d|\mathcal{P}(r)} \mathbb{P}E_d$$

$$= \sum_{d|\mathcal{P}(r)} \varphi^{-1}(d) \prod_{p|\mathcal{P}(r)/d} (1-\frac{1}{\varphi(p)}) \tag{3.7}$$

$$=1,$$

therefore $(K_D(x), \mathcal{A}, \mathbb{P})$ forms a probability space as well. According to (3.4), (3.5) and (3.6) we conclude that

$$|\nu A - \mathbb{P}A| \ll H \sum_{\substack{d\le z \\ d|\mathcal{P}(r)}} \varphi^{-1}(d) \prod_{p|\mathcal{P}(r)/d} (1-\frac{1}{\varphi(p)})$$

$$+ \pi^{-1}(\frac{x-1}{D}) \sum_{\substack{d\le z \\ d|\mathcal{P}(r)}} \sum_{\substack{t|\mathcal{P}(r)/d \\ t\le z^3}} 3^{\omega(t)} |R(x,td)| \tag{3.8}$$

$$+ \sum_{\substack{d>z \\ d|P(r)}} |\nu(E_d) - \mathbb{P}(E_d)|.$$

By (3.7) the first term on the right hand side above does not exceed H. To estimate the second term we note that

$$\sum_{t|n} 3^{\omega(t)} = 4^{\omega(n)},$$

therefore an application of the Cauchy-Schwarz inequality shows that

$$\begin{aligned}\sum_{\substack{d\leq z\\ d|\mathcal{P}(r)}} \sum_{\substack{t|\mathcal{P}(r)/d\\ t\leq z^3}} 3^{\omega(t)}|R(x,td)| &\leq \sum_{\substack{d\leq z^4\\ d|\mathcal{P}(r)}} |R(x,d)| \sum_{t|d} 3^{\omega(t)} \\ &\leq (\sum_{\substack{d\leq z^4\\ d|\mathcal{P}(r)}} \frac{8^{\omega(d)}}{\varphi(d)})^{1/2} (\sum_{\substack{d\leq z^4\\ d|\mathcal{P}(r)}} \varphi(d)|R(x,d)|^2)^{1/2} \\ &= \Pi_1 \cdot \Pi_2.\end{aligned}$$

Setting $z = x^{\sigma/12}$ we have that $z^4 \leq (\frac{x-1}{D})^{1/3}$, therefore the Brun-Tichmarsh Theorem is applicable and we deduce that

$$\begin{aligned}\Pi_2 &\leq \sqrt{\pi(\frac{x-1}{D})} (\sum_{\substack{d\leq z^4\\ d|\mathcal{P}(r)}} |R(x,d)|)^{1/2} \\ &\ll_A \pi(\frac{x-1}{D}) \log^{-A} x,\end{aligned}$$

where A is an arbitrary large positive constant. In this last step we used that the Bombieri-Vinogradov theorem is applicable as well. It remains to estimate Π_1^2. But it does not exceed

$$\begin{aligned}\sum_{\substack{d\\ P_1(d)\leq r}} \frac{\mu^2(d) 8^{\omega(d)}}{\varphi(d)} &= \prod_{p\leq r} (1 + \frac{8}{p-1}) \\ &\ll \log^8 x,\end{aligned}$$

therefore resetting A we obtain that thesecondn term on the right hand side of (3.8) is at most

$$\mathcal{O}(H + \log^{-A} x).$$

Since for all $0 < \gamma$

$$\begin{aligned}\sum_{\substack{d|\mathcal{P}_1(r)\\ d>z}} \frac{1}{d} &\ll \frac{1}{z^\gamma} \exp(\sum_{p\leq r} \frac{p^\gamma}{p}) \\ &\ll \frac{1}{z^\gamma} \exp(r^\gamma \log_2 r) \\ &\ll \log r \exp\{\exp(\gamma \log r) - \gamma \log z\},\end{aligned}$$

therefore by setting $\gamma = \frac{12}{\sigma} \log^{-1} r$ we have

$$\begin{aligned}
&\sum_{\substack{d>z \\ d|\mathcal{P}_1(r)}} \frac{1}{\varphi(d)} \prod_{p|\mathcal{P}_1(r)/d} (1 - \frac{1}{\varphi(p)}) \\
&\qquad = \prod_{p|\mathcal{P}_1(r)} (1 - \frac{1}{p}) \sum_{\substack{d>z \\ d|\mathcal{P}_1(r)}} \frac{1}{\varphi(d)} \prod_{p|\mathcal{P}_1(r)/d} (1 - \frac{1}{\varphi(p)}) \prod_{p|\mathcal{P}_1(r)} (1 - \frac{1}{p})^{-1} \\
&\qquad \ll \log^{-1} r \sum_{\substack{d>z \\ d|\mathcal{P}_1(r)}} \frac{1}{\varphi(d)} \prod_{p|d} (1 - \frac{1}{p})^{-1} \\
&\qquad = \log^{-1} r \sum_{\substack{d>z \\ d|\mathcal{P}_1(r)}} \frac{1}{d} \\
&\qquad \ll \exp(-\frac{\log x}{\log r}).
\end{aligned}$$

We deduced that

$$\sum_{\substack{d>z \\ d|\mathcal{P}(r)}} \mathbb{P}E_d \ll \exp(-\frac{\log x}{\log r}),$$

and that

$$\begin{aligned}
\sum_{\substack{d>z \\ d|\mathcal{P}(r)}} \nu E_d &= 1 - \sum_{\substack{d\leq z \\ d|\mathcal{P}(r)}} \nu E_d \\
&= 1 - \sum_{\substack{d\leq z \\ d|\mathcal{P}(r)}} \mathbb{P}E_d + \mathcal{O}(\exp(-\frac{\log x}{\log r}) + \log^{-A} x) \\
&= \sum_{\substack{d>z \\ d|\mathcal{P}(r)}} \mathbb{P}E_d + \mathcal{O}(\exp(-\frac{\log x}{\log r}) + \log^{-A} x) \\
&\ll \exp(-\frac{\log x}{\log r}) + \log^{-A} x,
\end{aligned}$$

therefore the third term on the right hand side of (3.8) does not exceed

$$\mathcal{O}(\exp(-\frac{\log x}{\log r}) + \log^{-A} x).$$

Putting it together we arrived at

$$|\nu A - \mathbb{P}A| \ll \mathcal{O}(\exp(-\frac{\log x}{\log r}) + \log^{-A} x)$$

uniformly for all $A \in \mathcal{A}$.

For all $p \leq r$ we define random variables X_p on $(K_D(x), \mathcal{A}, \mathbb{P})$ by

$$X_p(m) = \begin{cases} f(p) & \text{if } p|m \\ 0 & \text{otherwise .} \end{cases}$$

It is easy to see that X_p assumes the value $f(p)$ with probability $1/\varphi(p)$, and that for different values of p these are independent with respect to themeasuree $\mathbb{P}$.Furthermoree since for all $m \in K_D(x)$ we have that

$$\sum_{\substack{p \leq r \\ p \nmid D}} X_p(m) = f_r(m),$$

therefore the Lemma follows for all $r \leq z^{1/9}$. Noting that the estimation (3.3) remains true in an obvious way if $r > z^{1/9}$, the proof is ready. $\square$

Chapter 4

Distribution of additive arithmetic functions

In the last Chapter we discussed the pure probabilistic description of the distribution of additive functions. With this method we were able to prove the sufficiency of the existence of the distribution, and in this Chapter we want to get necessity. To characterize the additive functions for which a limit law exists, Kubilius's proof was based on the multiplicative structure of $\mathbb{N}$ in [26], i.e. the Euler product representation of the corresponding Dirichlet generating function. In our case the subsets of $\mathbb{N}$ obeys a rather complicated structure, therefore the Dirichlet generating function is not so regular.

In their original work [12], Erdős and Wintner used a method, which was a precursor to the Turán-Kubilius inequality, which will be useful for our purposes.

Lemma 18 (Turán-Kubilius). *Let $f(n)$ be a real additive function, and let*

$$E(x) := \sum_{p^k \le x} \frac{f(p^k)}{p^k} \left(1 - \frac{1}{p}\right),$$

$$D(x) := \left(\sum_{p^k \le x} \frac{|f(p^k)|^2}{p^k} \right)^{1/2}.$$

Then

$$\sum_{n \le x} |f(n) - E(x)|^2 \ll x D^2(x)$$

holds uniformly for all f and positive x.

(See [26], [25], [44], [45], [46].)

This inequality has been generalized in some sense for some rare subsets of the natural numbers like $\mathcal{P}+1$ ([8]). In this Chapter we will prove the cooresponding inequality for $\mathcal{P}_k+1$. This will be one of the main tools through the proofs.

The aim of this Chapter is to prove the following theorem

Theorem 1. *Let f be a real additive function. Let*

$$F_{k,x}(z) := \nu_x(n \in \mathcal{P}_k + 1; f(n) \le z).$$

Assume that there is a sequence $k = k_x$, for which all the terms have the property $A(\varepsilon, x)$, and a distribution function F such that $F_{k,x} \Rightarrow F$. Then the Erdős-Wintner condition holds.

Conversely, assume that the Erdős-Wintner condition holds for f. Then with a distribution function G we have

$$\max_{2\le k\le \varepsilon(x)\sqrt{\log\log x}} |F_{k,x}(z) - G(z)| \to 0 \quad (x \to \infty)$$

for all z continuity points of G. Consequently $F = G$.

The characteristic function of F is given by

$$\varphi(t) = \prod_p (1 + h(p)),$$

where

$$h(p) = -\frac{1}{p-1} + \sum_{m=1}^{\infty} \frac{e^{itf(p^m)}}{p^m}. \tag{4.1}$$

To prove this result we use a generalization of the result of Kátai for $D\mathcal{P}+1$, which is the following

Theorem 2. *Let f be a real additive function, and assume that*

$$\sum_{|f(p)|>1} \frac{1}{p}, \qquad \sum_{|f(p)|\le 1} \frac{f^2(p)}{p} \tag{4.2}$$

are convergent. Let $1 \ge \sigma > 0$, and $\varrho = \min\{\sigma/4, 1/4\}$. Let

$$A(x) := \sum_{\substack{|f(p)|\le 1\\ p\le x}} \frac{f(p)}{p}, \tag{4.3}$$

$$a(m) := \sum_{\substack{|f(p)|\le 1\\ p\mid m}} \frac{f(p)}{p}. \tag{4.4}$$

Let $K_D(x) = \{Dp + 1 \le x \mid p \in \mathcal{P}\}$.

$F_{D,x}(y) :=$

$$\nu_x\left(n \in K_D(x), f(n) - \left(A\left(\left(\frac{x-1}{D}\right)^{\varrho}\right) - a(D)\right) \le z\right), \tag{4.5}$$

with characteristic function $\varphi_{D,x}(t)$*, and let*

$$\varphi_D(t) := \prod_{\substack{|f(p)|>1 \\ p \nmid D}} (1 + h(p)) \prod_{\substack{|f(p)| \le 1 \\ p \nmid D}} (1 + h(p))\, e^{-it\frac{f(p)}{p}}. \tag{4.6}$$

Then

$$\max_{1 \le D \le x^{1-\sigma}} |\varphi_{D,x}(t) - \varphi_D(t)| \to 0 \quad (x \to \infty), \tag{4.7}$$

uniformly for all bounded values of t*, i.e. if* $|t| < T$*,* T *is an arbitrary constant.*

We start with the proof of this last result.

4.1 Proof of Theorem 2

First we show that the truncated additive function possess a limit law with characteristic function $\varphi_D(t)$, and then that the tail of the function contributes a negligible amount to the characteristic function.

Let y_x be tending to infinity so slowly that $y_x \le \frac{1}{2} \log\log\log\log x$, and let $x \gg_\sigma 1$, say. Let us define the additive function

$$f_0(p^\alpha) = \begin{cases} f(p^\alpha) & \text{if } p^\alpha \le y_x,\ \alpha \ge 1 \\ 0 & \text{otherwise,} \end{cases}$$

and consider the following distribution function:

$$G_{x,D}(z) := \nu_x\left(n \in K_D(x), f_0(n) - (A(y_x) - a(D)) \le z\right),$$

where $D \le x^{1-\sigma}$. Then the characteristic function of $G_{x,D}$ is

$$\psi_{x,D}(t) = \frac{1}{\# K_D(x)} e^{-it(A(y_x) - a(D))} \sum_{n \in K_D(x)} e^{itf_0(n)}. \tag{4.8}$$

We furthermore define g_x by

$$g_x = \mu * e^{itf_0}, \tag{4.9}$$

where μ is the Möbius function, and $*$ is the Dirichlet convolution. It is clear that $g_x(n)$ is multiplicative and that $g_x(p^\alpha) = e^{itf_0(p^\alpha)} - e^{itf_0(p^{\alpha-1})}$, and by the inequality $\pi(y_x) < 2\frac{y_x}{\log y_x}$ we get that $g_x(n) = 0$ for $n > e^{2y_x}$, i.e. if $n > \log\log\log x$. These together imply that

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{g_x(n)}{\varphi(n)} &= \prod_{p \le y_x} \left(1 + \sum_{\alpha \ge 1} \frac{e^{itf_0(p^\alpha)} - e^{itf_0(p^{\alpha-1})}}{p^{\alpha-1}(p-1)} \right) \\ &= \prod_{p \le y_x} \left(1 - \frac{1}{p-1} + \sum_{\alpha \ge 1} \frac{e^{itf_0(p^\alpha)}}{p^\alpha} \right). \end{aligned} \tag{4.10}$$

If $(D, d) = 1$ then there is a unique l_d modulo d defined by

$$Dl_d + 1 \equiv 0 \pmod{d},$$

via the Möbius inversion formula we therefore have

$$\sum_{n \in K_D(x)} e^{itf_0(n)} = \sum_{\substack{d \le x \\ (D,d)=1}} g_x(d) \pi \left(\frac{x-1}{D}, d, l_d \right). \tag{4.11}$$

Since $g_x(d) = 0$ if $d > \sigma \log x$, we can apply the Siegel-Walfisz Theorem, whence the right hand side of (4.11) is

$$\pi \left(\frac{x-1}{D} \right) \sum_{\substack{d=1 \\ (D,d)=1}}^{\infty} \frac{g_x(d)}{\varphi(d)} + \mathcal{O} \left(\pi \left(\frac{x-1}{D} \right) e^{-c\sqrt{\log x}} \sum_{\substack{d=1 \\ (D,d)=1}}^{\infty} \frac{|g_x(d)|}{\varphi(d)} \right).$$

In the error term

$$\begin{aligned} \sum_{\substack{d=1 \\ (D,d)=1}}^{\infty} \frac{|g_x(d)|}{\varphi(d)} &\le \prod_{p \le y_x} \left(1 + \sum_{\alpha=1}^{\infty} \frac{|g_x(p^\alpha)|}{\varphi(p^\alpha)} \right) \\ &\le \prod_{p \le y_x} \left(1 + \frac{2p}{(p-1)^2} \right) \\ &\ll \log\log\log x, \end{aligned}$$

say. Furthermore using the inequality

$$\left| \prod_1^n a_i - \prod_1^n b_i \right| \le \sum_1^n |a_i - b_i|,$$

which holds for all complex numbers a_i, b_i with $|a_i| \leq 1$, $|b_i| \leq 1$ we obtain that

$$\sum_{\substack{d=1\\(d,D)=1}}^{\infty} \frac{g_x(d)}{\varphi(d)} = \prod_{\substack{p\leq y_x\\p\nmid D}} \left(1 - \frac{1}{p-1} + \sum_{\alpha\geq 1} \frac{e^{itf_0(p^\alpha)}}{p^\alpha}\right)$$
$$= \prod_{\substack{p\leq y_x\\p\nmid D}} (1 + h(p)) + o(1) \quad (x \to \infty),$$

where the implied constant in $o(1)$ is absolute, it does not depend on t, and $h(p)$ is defined by (4.1). Thus we obtain, that

$$\psi_{x,D}(t) = \prod_{\substack{p\leq y_x\\|f(p)|>1\\p\nmid D}} (1 + h(p)) \prod_{\substack{p\leq y_x\\|f(p)|\leq 1\\p\nmid D}} (1 + h(p))e^{-it\frac{f(p)}{p}} + o\left(1\right)$$
$$(x \to \infty)$$

uniformly in $t \in \mathbb{R}$, and the constant implied by $o(1)$ is absolute.

Since using the convergence of (4.2)

$$\sum_{\substack{p\geq y_x\\p\nmid D\\|f(p)|>1}} |\log(1 + h(p))| \ll \sum_{\substack{p\geq y_x\\|f(p)|>1}} \frac{|e^{itf(p)} - 1|}{p} + \sum_{p\geq y_x} \frac{1}{p^2}$$
$$= o(1) \quad (x \to \infty),$$

and

$$\sum_{\substack{p\geq y_x\\p\nmid D\\|f(p)|\leq 1}} \left|\log(1 + h(p)) - it\frac{f(p)}{p}\right| = \sum_{\substack{p\geq y_x\\p\nmid D\\|f(p)|\leq 1}} \frac{\left|e^{itf(p)} - 1 - itf(p)\right|}{p}$$
$$+ \mathcal{O}\left(\sum_{p\geq y_x} \frac{1}{p^2}\right)$$
$$\ll t^2 \sum_{\substack{p\geq y_x\\p\nmid D\\|f(p)|\leq 1}} \frac{f^2(p)}{p^2} + \frac{1}{y_x}$$
$$= o(1) \quad (x \to \infty)$$

uniformly for all $|t| < T$, we deduce that

$$\max_{1\leq D\leq x^{1-\sigma}} |\psi_{x,D}(t) - \varphi_D(t)| \to 0 \quad (x \to \infty) \tag{4.12}$$

with the same uniformity for all $|t| < T$, where $\varphi_D(t)$ is given by (4.6).

Using Lévy 's continuity theorem it immediately follows, that $G_{D,x} \Rightarrow F_D$ for all fix D.

Next we define another two additive functions as follows: let

$$f_1(p^\alpha) = \begin{cases} f(p) & \text{if } y_x < p \le \left(\frac{x-1}{D}\right)^\varrho, \ \alpha = 1 \text{ and } |f(p)| \le 1 \\ 0 & \text{otherwise,} \end{cases}$$

and let

$$f_2(p^\alpha) = \begin{cases} f(p) & \text{if } \left(\frac{x-1}{D}\right)^\varrho < p \le \left(\frac{x-1}{D}\right)^{1-\vartheta_x}, \ \alpha = 1 \text{ and } |f(p)| \le 1 \\ 0 & \text{otherwise,} \end{cases}$$

where $\varrho < \min(\sigma/4, 1/4)$ and ϑ_x is tending to zero slowly.

With this functions we have that

$$\nu_x(n \in K_D(x), f(n) \ne f_0(n) + f_1(n) + f_2(n)) = o(1)\pi\left(\frac{x-1}{D}\right) \quad (x \to \infty), \tag{4.13}$$

with an appropriate ϑ_x.

To obtain this last identity, by using the notations

$$B = \{q \in \mathcal{P} \mid |f(q)| > 1\},$$

and

$$B_y = \{q \in \mathcal{P} \mid q > y, \ q \in B\}$$

it suffice to prove that

$$\#\{n \in K_D(x) \mid \exists q|n, \ q \in B_{y_x}\} = \delta(y_x)\pi\left(\frac{x-1}{D}\right), \tag{4.14}$$

$\delta(y_x) \to 0$ $(x \to \infty)$, and

$$\nu_x\left(n \in K_D(x); \exists q|n \ , \ q > \left(\frac{x-1}{D}\right)^{1-\vartheta_x}\right) = o(1)\pi\left(\frac{x-1}{D}\right) \quad (x \to \infty), \tag{4.15}$$

and

$$\#\{n \in K_D(x) \mid \exists q^2|n,\ q > y\} = o(1)\pi\left(\frac{x-1}{D}\right) \tag{4.16}$$
$$(x \to \infty).$$

To estimate the left hand side of (4.14) we split the corresponding set into three, not necessarily distinct parts. In the first part we take numbers which have a prime divisor q with $y_x < q \leq \left(\frac{x-1}{D}\right)^{\varrho}$. The second part contains the numbers, which have a prime divisor q such that $\left(\frac{x-1}{D}\right)^{\varrho} < q \leq \left(\frac{x-1}{D}\right)^{1-\vartheta_x}$. The third part contains the rest of the integers in $K_D(x)$. Let us denote the number of integers in this three parts by Σ_1 and Σ_2, Σ_3 respectively.

We find that

$$\Sigma_1 \leq \sum_{Dp+1 \leq x} \sum_{\substack{q|Dp+1 \\ y_x < q \leq \left(\frac{x-1}{D}\right)^{\varrho} \\ |f(q)|>1}} 1 \leq \sum_{\substack{y_x < q \leq \left(\frac{x-1}{D}\right)^{\varrho} \\ |f(q)|>1}} \pi\left(\frac{x-1}{D}, q, l_q\right),$$

and similarly

$$\Sigma_2 \leq \sum_{\substack{\left(\frac{x-1}{D}\right)^{\varrho} < q \leq \left(\frac{x-1}{D}\right)^{1-\vartheta_x} \\ |f(q)|>1}} \pi\left(\frac{x-1}{D}, q, l_q\right).$$

Using the Brun-Titchmarsh theorem for Σ_1 we obtain,

$$\Sigma_1 \leq c\pi\left(\frac{x-1}{D}\right) \sum_{\substack{y_x < q \\ |f(q)|>1}} \frac{1}{q}.$$

Using sieve estimates for Σ_2 we deduce,

$$\Sigma_2 \leq c \sum_{\substack{\left(\frac{x-1}{D}\right)^{\varrho} < q \leq \left(\frac{x-1}{D}\right)^{1-\vartheta_x} \\ |f(q)|>1}} \frac{x-1}{qD\log\frac{x-1}{qD}} \leq c\frac{1}{\vartheta_x}\frac{x-1}{D\log\left(\frac{x-1}{D}\right)} \sum_{\substack{\left(\frac{x-1}{D}\right)^{\varrho} < q \\ |f(q)|>1}} \frac{1}{q}.$$

With the choice

$$\vartheta_x^2 = \max\left(\sum_{\substack{\left(\frac{x-1}{D}\right)^{\varrho} < q \\ |f(q)|>1}} \frac{1}{q},\ \sum_{\substack{\left(\frac{x-1}{D}\right)^{\varrho} < q \\ |f(q)|\leq 1}} \frac{f^2(q)}{q}\right), \tag{4.17}$$

we arrive at $\Sigma_2 = o(1)\pi\left(\frac{x-1}{D}\right)$ as $x \to \infty$. We need the second expression in the definition of ϑ_x lather.

The estimation of Σ_3 coincides with that of (4.15). The corresponding quantity does not exceed

$$\sum_{Dp+1\le x} \sum_{\substack{Dp+1=aq \\ \left(\frac{x-1}{D}\right)^{1-\vartheta_x}<q}} 1 \le \sum_{a\le\left(\frac{x-1}{D}\right)^{\vartheta_x}} \sum_{\substack{Dp+1=aq \\ p\le\frac{x-1}{D}}} 1.$$

To estimate the inner sum on the right hand side, we note that if

$$Dp + 1 = aq$$

with p, q primes, then there are unique integers $l_D \pmod a$ and $c_D \pmod D$ such that

$$Dl_D = -1 + c_D a,$$

and

$$p = l_D + at, \quad q = c_D + Dt, \quad t \le \frac{x-1}{aD}.$$

Therefore, using Lemma 15 we have that Σ_3 is at most

$$c \sum_{a<\left(\frac{x-1}{D}\right)^{\vartheta_x}} \frac{x}{aD\log^2\frac{x}{aD}} \le c\frac{x}{D\log^2\frac{x}{D}} \sum_{a<\left(\frac{x-1}{D}\right)^{\vartheta_x}} \frac{1}{a},$$

which is $o(1)\pi\left(\frac{x-1}{D}\right)$ as $x \to \infty$.

Similarly the left hand side of (4.16) is for all $0 < a < 1$ at most

$$\sum_{y<q<\left(\frac{x-1}{D}\right)^a} \pi\left(\frac{x-1}{D}, q^2, l_q\right) + \frac{x-1}{D} \sum_{q\ge\left(\frac{x-1}{D}\right)^a} \frac{1}{q^2} = \delta(y)\pi\left(\frac{x-1}{D}\right), \tag{4.18}$$

where $\delta(y) \to 0$ $(y \to \infty)$.

(4.14) and (4.15) and (4.16) imply (4.13).

Let

$$A_{D,y_x}\left(\left(\frac{x-1}{D}\right)^{\varrho}\right) = \sum_{\substack{y_x<p\le\left(\frac{x-1}{D}\right)^{\varrho} \\ |f_1(p)|\le 1 \\ p\nmid D}} \frac{f_1(p)}{p}.$$

In (4.13) we have seen that for almost all integers

$$f(n) = f_0(n) + f_1(n) + f_2(n)$$

holds. We must take care of the contribution of f_1 and f_2. In this direction we prove a Turán-Kubilius type inequality, namely that

$$\begin{aligned}\sum_{Dp+1\le x}\left|f_1\left(Dp+1\right) - A_{D,y_x}\left(\left(\frac{x-1}{D}\right)^{\varrho}\right)\right|^2 \\ \le c\pi\left(\frac{x-1}{D}\right)\sum_{\substack{y_x<p\\ |f(p)|\le 1}}\frac{f^2\left(p\right)}{p},\end{aligned} \tag{4.19}$$

and

$$\sum_{Dp+1\le x} f_2^2(Dp+1) \le c\pi\left(\frac{x-1}{D}\right)\sum_{\substack{\left(\frac{x-1}{D}\right)^{\varrho}<p\\ |f(p)|\le 1}}\frac{f^2\left(p\right)}{p}. \tag{4.20}$$

Proof of (4.19) Let

$$B^2\left(\left(\frac{x-1}{D}\right)^{\varrho}\right) := \sum_{\substack{y_x<q\le\left(\frac{x-1}{D}\right)^{\varrho}\\ q\nmid D}}\frac{f_1^2\left(q\right)}{q}.$$

The left hand side of (4.19) can be written as

$$S_1 - 2A_{D,y_x}\left(\left(\frac{x-1}{D}\right)^{\varrho}\right)S_2 + A_{D,y_x}^2\left(\left(\frac{x-1}{D}\right)^{\varrho}\right)\pi\left(\frac{x-1}{D}\right),$$

where

$$\begin{aligned}S_1 =& \sum_{Dp+1\le x}\left(\sum_{q|Dp+1} f_1\left(q\right)\right)^2 \\ =& \sum_{\substack{y_x<q\le\left(\frac{x-1}{D}\right)^{\varrho}\\ q\nmid D}} f_1^2\left(q\right)\pi\left(\frac{x-1}{D}, q, l_q\right) \\ &+ \sum_{\substack{y_x\le q\le\left(\frac{x-1}{D}\right)^{\varrho}\\ q\nmid D}}\ \sum_{\substack{y_x\le q'\le\left(\frac{x-1}{D}\right)^{\varrho}\\ q'\nmid D\\ q\ne q'}} f_1\left(q\right)f_1\left(q'\right)\pi\left(\frac{x-1}{D}, qq', l_{qq'}\right) \\ =&\Sigma_{11} + \Sigma_{12},\end{aligned}$$

and

$$S_2 = \sum_{\substack{y_x < q \le \left(\frac{x-1}{D}\right)^{\varrho} \\ q \nmid D}} f_1(q)\, \pi\left(\frac{x-1}{D}, q, l_q\right).$$

Since $\varrho < \sigma/4$, thus we can estimate Σ_{11} using the Brun-Titchmarsh theorem, and we obtain

$$\Sigma_{11} < cB^2 \left(\left(\frac{x-1}{D}\right)^{\varrho}\right) \pi\left(\frac{x-1}{D}\right).$$

Moreover Σ_{12} equals

$$\begin{aligned}
&\pi\left(\frac{x-1}{D}\right) \sum_{\substack{y_x \le q \le \left(\frac{x-1}{D}\right)^{\varrho} \\ q \nmid D}} \sum_{\substack{y_x \le q' \le \left(\frac{x-1}{D}\right)^{\varrho} \\ q' \nmid D \\ q \ne q'}} \frac{f_1(q)\, f_1(q')}{\varphi(qq')} \\
&+ \sum_{\substack{y_x \le q \le \left(\frac{x-1}{D}\right)^{\varrho} \\ q \nmid D}} \sum_{\substack{y_x \le q' \le \left(\frac{x-1}{D}\right)^{\varrho} \\ q' \nmid D \\ q \ne q'}} f_1(q)\, f_1(q') \left(\pi\left(\frac{x-1}{D}, qq', l_{qq'}\right) - \frac{\pi\left(\frac{x-1}{D}\right)}{\varphi(qq')}\right) \\
&= \zeta + E,
\end{aligned}$$

such that

$$\Sigma_{12} \le A^2_{D,y_x} \left(\left(\frac{x-1}{D}\right)^{\varrho}\right) \pi\left(\frac{x-1}{D}\right) + E.$$

An application of the Cauchy-Schwarz inequality shows that E^2 is at most

$$\begin{aligned}
&\sum_{\substack{y_x < q \le \left(\frac{x-1}{D}\right)^{\varrho} \\ q \nmid D}} \sum_{\substack{y_x < q' \le \left(\frac{x-1}{D}\right)^{\varrho} \\ q' \nmid D \\ q \ne q'}} \frac{f_1^2(q)\, f_1^2(q')}{\varphi(q)\, \varphi(q')} \\
\times &\sum_{\substack{y_x < q \le \left(\frac{x-1}{D}\right)^{\varrho} \\ q \nmid D}} \sum_{\substack{y_x < q' \le \left(\frac{x-1}{D}\right)^{\varrho} \\ q' \nmid D \\ q \ne q'}} \varphi(qq') \left| \frac{\pi\left(\frac{x-1}{D}\right)}{\varphi(qq')} - \pi\left(\frac{x-1}{D}, qq', l_{qq'}\right)\right|^2.
\end{aligned}$$

Since $qq' < \left(\frac{x-1}{D}\right)^{1/2}$, the Brun-Titchmarsh theorem is applicable, and we deduce that E is at most

$$B^2\left(\left(\frac{x-1}{D}\right)^{\varrho}\right)\left(\pi\left(\frac{x-1}{D}\right)\right)^{1/2}$$

$$\times\left(\sum_{\substack{y_x<q\le\left(\frac{x-1}{D}\right)^{\varrho}\\ q\nmid D}}\ \sum_{\substack{y_x<q'\le\left(\frac{x-1}{D}\right)^{\varrho}\\ q'\nmid D\\ q\neq q'}}\left|\frac{\pi\left(\frac{x-1}{D}\right)}{\varphi\left(qq'\right)}-\pi\left(\frac{x-1}{D},qq',l_{qq'}\right)\right|\right)^{1/2}.$$

The Bombieri-Vinogradov theorem is also applicable, and we conclude that

$$\begin{aligned}\Sigma_{12}\le &A_{D,y_x}^2\left(\left(\frac{x-1}{D}\right)^{\varrho}\right)\pi\left(\frac{x-1}{D}\right)\\ &+\mathcal{O}(1)B^2\left(\left(\frac{x-1}{D}\right)^{\varrho}\right)\pi\left(\frac{x-1}{D}\right)\log x^{-A},\end{aligned}$$

where A is an arbitrary big positive constant. This leads to

$$\begin{aligned}S_1 = &A_{D,y_x}^2\left(\left(\frac{x-1}{D}\right)^{\varrho}\right)\pi\left(\frac{x-1}{D}\right)\\ &+\mathcal{O}(1)B^2\left(\left(\frac{x-1}{D}\right)^{\varrho}\right)\pi\left(\frac{x-1}{D}\right).\end{aligned}$$

To estimate S_2 we note, that using the Cauchy- Schwarz inequality we have

$$\begin{aligned}A_{D,y_x}\left(\left(\frac{x-1}{D}\right)^{\varrho}\right) &= \sum_{\substack{y_x<p\le\left(\frac{x-1}{D}\right)^{\varrho}\\ p\nmid D}}\frac{f_1(p)}{p}\\ &\le\left(\sum_{\substack{y_x<p\le\left(\frac{x-1}{D}\right)^{\varrho}\\ p\nmid D}}\frac{f_1^2(p)}{p}\sum_{\substack{y_x<p\le\left(\frac{x-1}{D}\right)^{\varrho}\\ p\nmid D}}\frac{1}{p}\right)^{1/2}\\ &\ll B\left(\left(\frac{x-1}{D}\right)^{\varrho}\right)(\log\log x)^{1/2},\end{aligned}$$

such that using the above method one can easily see, that

$$\begin{aligned}S_2 = &2A_{D,y_x}^2\left(\left(\frac{x-1}{D}\right)^{\varrho}\right)\pi\left(\frac{x-1}{D}\right)\\ &+o\left(1\right)B^2\left(\left(\frac{x-1}{D}\right)^{\varrho}\right)\pi\left(\frac{x-1}{D}\right)\end{aligned}$$

as $x\to\infty$, which implies (4.19).

Proof of (4.20) Since a positive integer $n\le x$ can only have a bounded number of distinct prime divisors $q>\left(\frac{x-1}{D}\right)^{\varrho}$, we have that (4.20) does not exceed

$$c\sum_{\left(\frac{x-1}{D}\right)^{\varrho}<q\le\left(\frac{x-1}{D}\right)^{1-\vartheta_x}} f_2^2(q)\pi\left(\frac{x-1}{D},q,l_q\right).$$

Using Lemma 2 this is at most

$$c\frac{1}{\vartheta_x}\frac{x-1}{D\log\frac{x-1}{D}}\sum_{\left(\frac{x-1}{D}\right)^{\varrho}<q\le\left(\frac{x-1}{D}\right)^{1-\vartheta_x}}\frac{f_2^2(q)}{q},$$

which thanks to the choice (4.17) implies (4.20).

Using (4.19) and (4.20) we obtain that

$$\begin{aligned}
&\frac{1}{K_D(x)}\sum_{n\in K_D(x)}\left|e^{it\left(f_0(n)+f_1(n)+f_2(n)-A\left(\left(\frac{x-1}{D}\right)^{\varrho}\right)-a(D)\right)}-e^{it(f_0(n)-(A(y_x)-a(D))}\right|^2\\
&\ll\frac{1}{K_D(x)}t^2\sum_{n\in K_D(x)}\left|f_1(n)-A_{D,y_x}\left(\left(\frac{x-1}{D}\right)^{\varrho}\right)\right|^2\\
&\qquad+\frac{1}{K_D(x)}t^2\sum_{n\in K_D(x)}f_2^2(n)\\
&=o(1),\quad (x\to\infty)
\end{aligned}$$

uniformly for all $|t|<T$, and $1\le D\le x^{1-\sigma}$.

Using this and (4.12) we have proved that

$$\sup_{D\le x^{1-\sigma}}\left|\frac{1}{\#K_D(x)}e^{-it\left(A\left(\left(\frac{x-1}{D}\right)^{\varrho}\right)-a(D)\right)}\sum_{n\in K_D(x)}e^{itf(n)}-\varphi_D(t)\right|\to 0\quad (x\to\infty),$$

uniformly as $|t|\le T$, T is an arbitrary constant.

4.2 Proof of Theorem 1.

4.2.1 Concluding the sufficiency part of Theorem 1. from Theorem 2.

Consider first the following sequence of distribution functions:

$$F_{k,x}(z)=\nu_x\left(n\in U_k(x)\ :\ f(n+1)-A(x)\le z\right),$$

where

$$A(x) = \sum_{\substack{p\le x\\ |f(p)|\le 1}} \frac{f(p)}{p}.$$

With the notations of Lemma 9., an application of Lemma 8 shows that the characteristic function ψ_x of F_x is

$$\frac{1}{B_k(x)} e^{-itA(x)} \sum_{\pi_{k-1}\in S_x} \sum_{p^*(\pi_{k-1})<p\le \frac{x}{\pi_{k-1}}} e^{itf(\pi_{k-1}p+1)}.$$

Since $A(x)$ is convergent, therefore

$$e^{itA(x)} = e^{itA_{\pi_{k-1},y_x}((\frac{x-1}{\pi_{k-1}})^{\varrho})} + o(1) \quad (x\to\infty)$$

uniformly for all $\pi_{k-1}\in S_x$. Using Theorem 2. and Lemma 8 this leads to

$$\begin{aligned}\psi_x(t) = \frac{1}{B_k(x)} &\sum_{\pi_{k-1}\in S_x} \pi\left(\frac{x}{\pi_{k-1}}\right)\left(\varphi_{\pi_{k-1}}(t)+o(1)\right)\\ &- \frac{1}{B_k(x)} \sum_{\pi_{k-1}\in S_x} \pi\left(p^*(\pi_{k-1})\right)\left(\varphi_{\pi_{k-1}}(t)+o(1)\right),\end{aligned} \tag{4.21}$$

for all $|t|<T$, as $x\to\infty$. As a corollary of the remark after the proof of Lemma 9 we obtain that $p^*(\pi_{k-1})\le x^{1/2}$ in the second term, which holds for all $\pi_{k-1}\in S_x$ and we get that this is $o(1)$ as $(x\to\infty)$. Now using the identity

$$\varphi(t) = \varphi_D(t)\,K_D(t),$$

where

$$K_D(t) = \prod_{\substack{p|D\\ |f(p)|>1}} (1+h(p)) \prod_{\substack{p|D\\ |f(p)|\le 1}} (1+h(p))\, e^{-it\frac{f(p)}{p}}$$

we obtain that the main term in (4.21) is:

$$\begin{aligned}\Sigma_1 = \varphi(t)\frac{1}{B_k(x)} &\sum_{\pi_{k-1}\in S_x} \pi\left(\frac{x}{\pi_{k-1}}\right)\\ &+ \frac{1}{B_k(x)} \sum_{\pi_{k-1}\in S_x} \pi\left(\frac{x}{\pi_{k-1}}\right)\varphi_{\pi_{k-1}}(t)\left(1-K_{\pi_{k-1}}(t)\right)\\ &+ o(1)\frac{1}{B_k(x)} \sum_{\pi_{k-1}\in S_x} \pi\left(\frac{x}{\pi_{k-1}}\right)\end{aligned} \tag{4.22}$$

as $(x \to \infty)$.

Arguing again with the remark after Lemma 9 we have that

$$\begin{aligned}\sum_{\pi_{k-1} \in S_x} \pi(p^*(\pi_{k-1})) &\ll x^{\frac{1}{2}+\frac{1}{A_x}} \\ &= B_k(x)o(1)\end{aligned}$$

as $(x \to \infty)$, therefore

$$\frac{1}{B_k(x)} \sum_{\pi_{k-1} \in S_x} \pi\left(\frac{x}{\pi_{k-1}}\right) \to 1, \quad (x \to \infty). \tag{4.23}$$

We get that

$$\Sigma_1 = \varphi(t)(1 + o(1)), \quad (x \to \infty)$$

uniformly for all $|t| \le T$. To see this last identity we calculate the second term in (4.22). We have

$$\log K_{\pi_{k-1}}(t) = \sum_{\substack{p|\pi_{k-1} \\ |f(p)|>1}} \log(1 + h(p)) + \sum_{\substack{p|\pi_{k-1} \\ |f(p)|\le 1}} \log(1 + h(p)) - \frac{itf(p)}{p}.$$

If $p|\pi_{k-1}$, then $p > M_x$, therefore if x is large enough then the first term here is

$$\begin{aligned}\sum_{\substack{p|\pi_{k-1} \\ |f(p)|>1}} \frac{e^{itf(p)} - 1}{p} + \mathcal{O}\left(\sum_{\substack{p|\pi_{k-1} \\ |f(p)|>1}} \frac{1}{p^2}\right) &= \mathcal{O}\left(\sum_{\substack{p \ge M_x \\ |f(p)|>1}} \frac{1}{p}\right) + \mathcal{O}\left(\sum_{\substack{p \ge M_x \\ |f(p)|>1}} \frac{1}{p^2}\right) \\ &= o(1)\end{aligned}$$

as $(x \to \infty)$, whilst the second term is

$$\begin{aligned}&\sum_{\substack{p|\pi_{k-1} \\ |f(p)|\le 1}} \frac{e^{itf(p)} - 1 - itf(p)}{p} + \mathcal{O}\left(\sum_{\substack{p|\pi_{k-1} \\ |f(p)|\le 1}} \frac{1}{p^2}\right) \\ &= |t|^2 \mathcal{O}\left(\sum_{\substack{p \ge M_x \\ |f(p)|\le 1}} \frac{f^2(p)}{p}\right) + \mathcal{O}\left(\sum_{\substack{p \ge M_x \\ |f(p)|\le 1}} \frac{1}{p^2}\right),\end{aligned}$$

therefore

$$K_{\pi_{k-1}}(t) - 1 = o(1) \quad (x \to \infty),$$

for all $|t| < T$, which together with the convergence of $A(x)$ implies our assertion.

4.2.2 Proof of the necessity part of Theorem 1.

In this section we use some ideas appeared in [21].

Lemma 19. *Let f be an additive function. Assume that $f(n) = c\log n + g(n)$, and that the series*

$$\sum_{|g(p)|>1} \frac{1}{p-1}, \qquad \sum_{|g(p)|\le 1} \frac{g^2(p)}{p-1}$$

converge. Define

$$U(x) = c\log x + \sum_{\substack{p\le x\\ |g(p)|\le 1}} \frac{g(p)}{p-1}.$$

Then $\nu_x(n\in\mathcal{P}_k,\, n\le x:\ f(n+1) - U(x) \le z)$ converges weakly to a limit distribution as $x\to\infty$. The characteristic function of the limit law is

$$\chi(t) := \frac{1}{1+itc} \prod_{|g(p)|>1} \left(1 - \frac{1}{p-1} + \sum_{m\ge 1} \frac{e^{itg(p^m)}}{p^m}\right) \times \prod_{|g(p)|\le 1} \left(1 - \frac{1}{p-1} + \sum_{m\ge 1} \frac{e^{itg(p^m)}}{p^m}\right) e^{-it\frac{g(p)}{p}},$$

and the limit distribution is continuous if and only if

$$\sum_{f(p)\ne 0} \frac{1}{p}$$

diverges.

Proof. Let

$$A(x) = \sum_{\substack{p\le x\\ |g(p)|\le 1}} \frac{g(p)}{p-1}.$$

Then this last lemma shows, that the distributions

$$\nu_x(n\in\mathcal{P}_k,\, n\le x:\ g(n+1) - A(x) \le z)$$

possess a limit law with characteristic function $\xi(t)$, say. Let

$$\varphi_x(t) = \frac{1}{\pi_k(x)} \sum_{\pi_k\le x} e^{itg(\pi_k+1)}.$$

We have

$$\varphi_x(t)\, e^{-itA(x)} \to \xi(t) \quad (x \to \infty).$$

Consider next the following sum

$$\frac{1}{\pi_k(x)} \sum_{\pi_k \le x} e^{itf(\pi_k+1)} = \frac{1}{\pi_k(x)} \sum_{\pi_k \le x} (\pi_k+1)^{itc}\, e^{itg(\pi_k+1)} \tag{4.24}$$

Remember that k may depend on x. Let us introduce the following notation

$$\Psi_{k,x}(y,t) := \sum_{\pi_{k_x} \le y} e^{itg(\pi_k+1)}$$

Applying Abel's summation formula for the right hand side of (4.24) we get that it is

$$\varphi_x(t)\, x^{itc} - itc \frac{1}{\pi_k(x)} \int_1^x y^{itc-1} \Psi_{k,x}(y,t)\, dy. \tag{4.25}$$

Since for $x^\lambda < y < x$ using Lemma 7 we have

$$\lim_{x\to\infty} \frac{1}{\pi_{k_x}(y)} \Psi_{k,x}(y,t) e^{-itA(y)} = \xi(t),$$

thus for this y

$$\begin{aligned}\Psi_{k,x}(y,t) = \xi(t)\, e^{itA(x)} + \xi(t)\, e^{itA(y)} \left\{1 - e^{it(A(x)-A(y))}\right\} \\ + o(1) \quad (x \to \infty).\end{aligned}$$

Since

$$A(x) - A(y) = o(1) \quad (x \to \infty)$$

if $x^\lambda < y < x$, we have

$$\frac{1}{\pi_{k_x}(y)} \Psi_{k,x}(y,t) = \xi(t)\, e^{itA(x)} + o(1) \quad (x \to \infty). \tag{4.26}$$

Using this we have that the integral in (4.25) equals

$$\begin{aligned}\int_1^{x^\lambda} y^{itc-1} \Psi_{k,x}(y,t)\, dy \;+\; \xi(t)\, e^{itA(x)} \int_{x^\lambda}^x y^{itc-1} \pi_{k_x}(y)\; dy \\ +\; o(x) \quad (x \to \infty).\end{aligned} \tag{4.27}$$

The first term is

$$\mathcal{O}(x^\lambda) = \pi_k(x) o(1) \quad (x \to \infty).$$

Using Lemma 7 we get that the second term in (4.27) is

$$\xi(t) e^{itA(x)} \int_{x^\lambda}^{x} y^{itc-1} \left(\frac{y}{\log y} \frac{\log \log^{k-1} y}{(k-1)!} (1 + o(1)) \right) dy,$$

which is $\frac{1}{(itc+1)} \xi(t) e^{itA(x)} \pi_k(x) + \pi_k(x) o(1)$. Here we used, that k depends only upon the upper bound of the integration, and that

$$\left(\frac{x^{itc+1}}{\log x} \frac{\log \log^{k-1} x}{(k-1)!} \right)' = (itc+1) \frac{x^{itc}}{\log x} \frac{\log \log^{k-1} x}{(k-1)!} + \mathcal{O}\left(\frac{\log \log^{k-2} x}{\log^2 x} \frac{1}{(k-2)!} \right).$$

We had shown, that (4.25) equals

$$\xi(t) \frac{e^{itA(x)} x^{itc}}{1 + itc} + o(1) \quad (x \to \infty).$$

We get that

$$\frac{1}{\pi_k(x)} \sum_{\pi_k \le x} e^{it(f(\pi_k + 1))} = \frac{x^{ict} e^{itA(x)}}{1 + ict} \xi(t) + o(1) \quad (x \to \infty),$$

therefore

$$\frac{1}{\pi_k(x)} \sum_{\pi_k \le x} e^{it(f(\pi_k + 1) - U(x))} = \frac{\xi(t)}{1 + ict} + o(1) \quad (x \to \infty),$$

and our lemma immediately follows. □

Lemma 20 (Sieve estimate)**.** *Let q, q' be prime numbers, and $x > x_0$. We have*

(4.28)
$$A = \#\{n \le x : qn + 1, q'n + 1 \in U_k(x)\} \ll \frac{x}{\log^2 x} \left(\frac{(\log \log x)^{k-1}}{(k-1)!} \right)^2 \Psi(|q - q'|),$$

where

$$\Psi(n) = \prod_{\substack{p \mid n \\ p > 2}} \frac{p-1}{p-2}.$$

Proof. We express a typical element $\pi_k \in \mathcal{P}_k$ as $\pi_k = \pi_{k-1}p$, where $P(\pi_{k-1}) < p$. Using Lemma 8 we need only to count the elements of the set

$$A = \#\{n \leq x : qn+1 = \pi_{k-1}p,\ q'n+1 = \pi'_{k-1}p',\ \text{and } \pi_{k-1} \leq x^\beta,\ \pi'_{k-1} \leq x^\beta\}.$$

Recall that

$$p(\pi_{k-1}) \geq M_x, \qquad p(\pi'_{k-1}) \geq M_x.$$

We have

$$\begin{aligned} qn+1 &= \pi_{k-1}p \\ q'n+1 &= \pi'_{k-1}p' \end{aligned}$$

so by the Chinese remainder theorem

$$n \equiv l_{\pi_{k-1}\pi'_{k-1}} \pmod{\pi_{k-1}\pi'_{k-1}}$$

with a unique $l_{\pi_{k-1}\pi'_{k-1}}$. So $n = t\pi_{k-1}\pi'_{k-1} + l_{\pi_{k-1}\pi'_{k-1}}$, where $t \leq \frac{x}{\pi_{k-1}\pi'_{k-1}}$. For p, p' we have

$$\begin{aligned} p &= qt\pi'_{k-1} + \frac{ql+1}{\pi_{k-1}} \\ p' &= q't\pi_{k-1} + \frac{q'l+1}{\pi'_{k-1}} \end{aligned} \qquad t \leq \frac{x}{\pi_{k-1}\pi'_{k-1}}.$$

Using sieve estimates (Lemma 15) we have, that the number of such p, p' is not more than

(4.29)

$$\begin{aligned} A &\ll \sum_{\substack{\pi_{k-1} \leq x^\beta \\ \pi'_{k-1} \leq x^\beta}} \Psi\left(q\pi'_{k-1}q'\pi_{k-1}\left|q\pi'_{k-1}\frac{q'l+1}{\pi'_{k-1}} - q'\pi_{k-1}\frac{ql+1}{\pi_{k-1}}\right|\right)\frac{x}{\log^2 x}\frac{1}{\pi_{k-1}\pi'_{k-1}} \\ &\ll \Psi\left(|q-q'|\right)\frac{x}{\log^2 x}\sum_{\substack{\pi_{k-1} \leq x^\beta \\ \pi'_{k-1} \leq x^\beta}} \Psi\left(\pi'_{k-1}\pi_{k-1}\right)\frac{1}{\pi_{k-1}\pi'_{k-1}}. \end{aligned}$$

Here we observed, that

$$q\pi'_{k-1}\frac{q'l+1}{\pi'_{k-1}} - q'\pi_{k-1}\frac{ql+1}{\pi_{k-1}} = q - q'.$$

The sum in the last expression of (4.29) is

$$\sum_{\pi_{k-1}\le x^\beta}\prod_{p_i|\pi_{k-1}}\frac{p_i-1}{(p_i-2)p_i}\sum_{\pi'_{k-1}\le x^\beta}\prod_{p'_i|\pi'_{k-1}}\frac{p'_i-1}{(p'_i-2)p'_i}$$
$$\ll\sum_{\pi_{k-1}\le x^\beta}\prod_{p_i|\pi_{k-1}}\frac{1}{(p_i-2)}\sum_{\pi'_{k-1}\le x^\beta}\prod_{p'_i|\pi'_{k-1}}\frac{1}{(p'_i-2)}$$
$$\ll\left(\frac{1}{(k-1)!}\left(\sum_{p\le x^\beta}\frac{1}{p}+\sum_{p>M_x}\frac{1}{p^2}\right)^{k-1}\right)^2,$$

which implies our assertion. □

Lemma 21. *With suitable constant δ_1 and c_1, and multiplicative $g:\mathbb{N}\to\mathbb{C}$ such that $|g|=1$, and*

$$\max_{1\le l\le c_1}\left|\frac{1}{x}\sum_{n\le x}g(n)^l\right|\le\delta_1$$

we have

$$\left|\frac{1}{B_k(x)}\sum_{n\in U_k(x)}g(n+1)\right|\le 1-\delta_1. \tag{4.30}$$

Proof. It is enough to prove that if the conditions hold, then

$$1-\left|\frac{1}{B_k(x)}\sum_{\pi_k\in U_k(x)}g(\pi_k+1)\right|\gg 1.$$

Some computation shows that

$$1-\left|\frac{1}{B_k(x)}\sum_{\pi_k\in U_k(x)}g(\pi_k+1)\right|\ge\frac{1}{2B_k(x)}\sum_{\pi_k\in U_k(x)}|1-wg(\pi_k+1)|^2, \tag{4.31}$$

with an appropriate complex w, with absolute value 1. Setting

$$R(Q)=\frac{1}{B_k(x)}\sum_{\pi_k\in U_k(x)}|1-wg(\pi_k+1)|^2\sum_{\substack{q|\pi_k+1\\ Q<q<2Q}}1,$$

we get with $0 < \eta < 1/4$ and $x \geq 2^{1/\eta}$ after some computation that the right hand side of (4.31) is at least

$$\frac{\eta^2}{2\log 2}\log x \min_{x^\eta \leq Q \leq x^{2\eta}} R(Q),$$

so it is enough to prove that $R(Q) \gg \frac{1}{\log x}$ uniformly with $x > x_0$, $x^\eta \leq Q \leq x^{2\eta}$. Let $\delta < 1/4$. There is an ω complex number with absolute value 1 such that

$$\sum_{\substack{Q<q<2Q \\ |g(q)-\omega|\leq\delta}} 1 \geq \frac{\delta}{10}\frac{Q}{\log Q}. \tag{4.32}$$

Set

$$S = \sum_{\pi_k \in U_k(x)} \sum_{\substack{q|\pi_k+1 \\ Q<q<2Q}}{}' 1.$$

The $'$ in the inner sum means the restriction to prime numbers q for which $|g(q) - \omega| \leq \delta$ with this appropriate ω. We get a lower bound for this sum applying Lemma 10 and Lemma 12.

$$S = \sum_{Q<q<2Q}{}' B_k(x, q, -1) \geq \frac{B_k(x)}{4Q} \sum_{Q<q<2Q}{}' 1 + \mathcal{O}\left(\frac{B_k(x)}{\log^A x}\right).$$

If $\pi_k + 1 = nq$, and $q^2 \not| \pi_k + 1$, then after some computation we have

$$|1 - wg(\pi_k + 1)| \geq |g(n) - \overline{w\omega}| - |g(q) - \omega|.$$

Thus $|g(q) - \omega| \leq \delta$ implies $|g(n) - \overline{w\omega}| \leq 2\delta$ or $|1 - wg(\pi_k + 1)| > \delta$.

Let

$$S_1 = \sum_{\pi_k \in U_k(x)} \sum_{Q<q<2Q}{}' \sum_{\substack{n \\ qn=\pi_k+1}}{}'' 1,$$

where $''$ in the inner sum means the restriction to integers for which

$$|g(n) - \overline{w\omega}| \leq 2\delta.$$

Let

$$S_2 = \sum_{\pi_k \in U_k(x)}{}^{*} \sum_{\substack{q|\pi_k+1 \\ Q<q<2Q}}{}' 1,$$

where $*$ in the outer sum means the restriction to the prime numbers for which $|1 - wg(\pi_k + 1)| > \delta$. Let

$$S_3 = \sum_{\pi_k \in U_k(x)} \sideset{}{'}\sum_{\substack{q^2 | \pi_k + 1 \\ Q < q < 2Q}} 1,$$

We have $S \leq S_1 + S_2 + S_3$. It is easy to see that $S_3 \leq x^{1-\eta}$. We use $R(Q)$ to estimate S_2.

$$S_2 \leq \sum_{\pi_k \in U_k(x)} \frac{|1 - wg(\pi_k + 1)|^2}{\delta^2} \sum_{\substack{q | \pi_k + 1 \\ Q < q \leq 2Q}} 1 = \frac{B_k(x)}{\delta^2} R(Q).$$

Putting it all together it is enough to prove that

$$S_1 \leq \frac{B_k(x)}{8Q} \sideset{}{'}\sum_{Q < q < 2Q} 1.$$

We have

$$S_1 \leq \sideset{}{''}\sum_{n \leq \frac{x+1}{Q}} \sideset{}{'}\sum_{\substack{Q < q < 2Q \\ qn - 1 = \pi_k}} 1.$$

Applying the Cauchy-Schwarz inequality we get

$$(S_1)^2 \leq \underbrace{\left(\sideset{}{''}\sum_{n \leq \frac{x+1}{Q}} 1 \right)}_{S_{11}} \underbrace{\left\{ \sum_{n \leq \frac{x+1}{Q}} \left(\sideset{}{'}\sum_{\substack{Q < q < 2Q \\ qn - 1 = \pi_k}} 1 \right)^2 \right\}}_{S_{12}}.$$

We have

$$S_{12} = \sideset{}{''}\sum_{Q < q, q' < 2Q} \sideset{}{'}\sum_{\substack{n \leq \frac{x+1}{Q} \\ qn - 1 = \pi_k \\ q'n - 1 = \pi_k'}} 1.$$

If $q = q'$ then this sum is $\mathcal{O}(x)$. For the other case we can use Lemma 20., and we get

$$S_{12} \ll \frac{x}{Q \log^2 x} \left(\frac{(\log \log x)^{k-1}}{(k-1)!} \right)^2 \sideset{}{'}\sum_{Q < q < 2Q} \sideset{}{'}\sum_{Q < q' < q} \Psi(q - q') + x.$$

After some computation we get that the inner sum does not exceed

$$c\left(\frac{\delta}{10}\right)^{-1/2}\sideset{}{'}\sum_{Q<q<2Q}1$$

(see Hildebrand).

We have

$$S_{12}\ll\delta^{-1/2}\frac{x}{Q\log^2 x}\left(\frac{(\log\log x)^{k-1}}{(k-1)!}\right)^2\left(\sideset{}{'}\sum_{Q<q<2Q}1\right)^2+x.$$

It is not so difficult to proof that if the conditions of the lemma holds, then

$$S_{11}\leq\frac{x}{Q}\delta\log\frac{1}{\delta}.$$

Putting it together we get

$$S_1\ll\left(\delta^{1/5}\frac{\pi_k(x)}{Q}+\frac{x}{\delta}\frac{\log Q}{Q^{3/2}}\right)\sideset{}{'}\sum_{Q<q<2Q}1.$$

Choosing small δ, we finished the outline of the proof. $\square$

The following two lemmas can be found in [21], thus we give only remarks according to our case.

Lemma 22. *Assume that*

$$\min_{\substack{1\leq l\leq c_1\\ \tau\leq c_2}} Re\sum_{p\leq x}\frac{1-g(p)^l p^{-i\tau}}{p}>c_2,$$

with a suitable $c_2>0$. Then (4.30) holds.

Remark. One can prove that if this condition holds, then the condition of Lemma 21. holds too, with a suitable large c_2.

Lemma 23. *There are constants δ_3, c_3 such that for fixed $x\geq 2$, $h>0$ and*

$$\max_{a\in\mathbb{R}}\{n\in U_k(x)\,:\,f(n+1)\in[a,a+h]\}\geq(1-\delta_3)B_k(x)$$

we have

$$\min_{|\lambda|\leq c_3h^2}\sum_{p\leq x}\frac{(\min(h,f(p)-\lambda\log p))^2}{p}\leq c_3h^2.$$

Proof. We must show that the condition of the previous lemma is violated if $h = 1$. We have

$$\left|\frac{1}{B_k(x)} \sum_{\pi_k \in U_k(x)} e^{itf(\pi_k+1)}\right| = \left|\frac{1}{B_k(x)} \sum_{\pi_k \in U_k(x)} e^{it(f(\pi_k+1)-a)}\right| \geq 1 - |t| - 2\delta_3.$$

Set $\delta_3 \leq \frac{\delta_1}{4}$ and $|t| \leq \frac{\delta_1}{2}$. The result of Lemma 21 is violated. We have using Lemma 22 that for $k = k(t) \leq c_1$ and $\tau = \tau(t)$, $|\tau| \leq c_2$

$$\sum_{p \leq x} \frac{Re\left(1 - e^{itf\left(p^k\right)p^{-i\tau}}\right)}{p} \leq c_2.$$

□

The following lemma can be found in [8] Lemma 1.9

Lemma 24. *Let F_x be a sequence of distribution functions, and let α, β be real functions. If $F_x(z + \alpha(x))$, and $F_x(z + \beta(x))$ converge weakly to a limit distribution with $x \to \infty$, than $\lim_{x\to\infty} (\alpha(x) - \beta(x))$ exists, and is finite.*

Proof of the necessity part of Theorem 1.
Assume f has a limit distribution. Then

$$\#\{n \leq x,\, n \in \mathcal{P}_k :\, f(n+1) \in [-z, z]\} \geq (1 - \delta_3)\,\pi_k(x)$$

holds for some z real number. We get using Lemma 23 with $h = 2z$

$$\sum_{p \leq x} \frac{\left(\min\left(h, f(p) - \lambda_x \log p\right)\right)^2}{p} \leq c_3 h^2$$

with $|\lambda_x| \leq c_3 h^2$. It follows that

$$\sum_{p} \frac{\left(\min\left(h, f(p) - \lambda \log p\right)\right)^2}{p}$$

converge for some λ. We immediately arrive at $g(n) = f(n) - \lambda \log n$ the series

$$\sum_{|g(p)|>1} \frac{1}{p-1}, \qquad \sum_{|g(p)|\leq 1} \frac{g(p)^2}{p-1}$$

converge. From Lemma 19, and Lemma 24 we get that

$$\lambda \log x + \sum_{\substack{p \le x \\ |g(p)| \le 1}} \frac{g(p)}{p-1}$$

converge. This implies $\lambda = 0$, and the convergence of

$$\sum_{|f(p)| \le 1} \frac{f(p)}{p-1},$$

and the proof of Theorem 1 is completed. $\square$

Chapter 5

Mean value estimates for multiplicative functions

5.1 General estimations

In the distribution theory of additive functions the mean value estimates for unimodular arithmetic functions plays a crucial role. In Chapter 4 we have used the following estimate

$$
\frac{1}{\pi_k(x)} \sum_{\pi_k \le x} e^{itf(\pi_k+1)} = \prod_{p \le x} \left(1 - \frac{1}{p-1} + \sum_{m \ge 1} \frac{e^{itf(p^m)}}{p^m} \right) + o(1) \quad (x \to \infty), \tag{5.1}
$$

for all real bounded t, whenever the Erdős-Wintner condition holds for f. It is easy to see that if this condition holds then

$$
\sum_p \frac{e^{itf(p)} - 1}{p} \tag{5.2}
$$

converges. Indeed

$$
\begin{aligned}
\sum_p \frac{e^{itf(p)} - 1}{p} &= \sum_{\substack{p \\ |f(p)| \le 1}} \frac{e^{itf(p)} - 1}{p} + \sum_{\substack{p \\ |f(p)| > 1}} \frac{e^{itf(p)} - 1}{p} \\
&= t \sum_{\substack{p \\ |f(p)| \le 1}} \frac{f(p)}{p} + \mathcal{O}\left(t^2 \sum_{\substack{p \\ |f(p)| \le 1}} \frac{f(p)^2}{p} + \sum_{\substack{p \\ |f(p)| > 1}} \frac{1}{p} \right).
\end{aligned}
$$

We will see that for the existence of the mean value (5.1) it is enough to assume the convergence of (5.2).

In the middle of the twentieth century Delange did some pioneering work about mean value estimations for multiplicative functions. One of his result was the following (See [7])

Theorem (Delange)**.** *Let g be a multiplicative function with $|g(n)| \leq 1$, satisfying*

$$\sum_p \frac{1 - Reg(p)}{p} < \infty.$$

Then we have

$$\frac{1}{x}\sum_{n\leq x} g(n) = \prod_{p\leq x}\left(1-\frac{1}{p}\right)\prod_{p\leq x}\left(1+\sum_{m\geq 1}\frac{g(p^m)}{p^m}\right) + o(1)$$

as x tends to infinity.

Although this result provides sufficient condition for multiplicative functions to have zero mean value, the full description of such multiplicative functions was given by Wirsing([48]) for real and by Halász ([17]) for complex multiplicative functions of modulus ≤ 1. The result of Halász extends Delange's theorem in the following way:

Theorem (Halász)**.** *Let g be a multiplicative function with $|g(n)| \leq 1$, satisfying*

$$\sum_p \frac{1 - Reg(p)p^{-i\tau}}{p} < \infty$$

for some real τ. Then we have

$$\frac{1}{x}\sum_{n\leq x} g(n) = \frac{x^{i\tau}}{1+i\tau}\prod_{p\leq x}\left(1-\frac{1}{p}\right)\prod_{p\leq x}\left(1+\sum_{m\geq 1}\frac{g(p^m)}{p^{m(1+i\tau)}}\right) + o(1)$$

as x tends to infinity. On the other hand, if there is no such τ then

$$\frac{1}{x}\sum_{n\leq x} g(n) = o(1) \quad (x \to \infty).$$

For an elementary proof see for example in [5] as well as in [23]. Although in his whole generality this result does not apply to multiplicative functions outside the unit disk, by making several restrictions on the growth of the function on the set of primes it is still possible to get asymptotic. (See [47])

More generally one can ask if a multiplicative function g, $|g(n)| \le 1$ has mean value in the set $\mathcal{P}_k + 1$. For $k = 1$ it was proved by Kátai [29] that for this purpose it is enough that

$$\sum_p \frac{g(p) - 1}{p}$$

converges. In this case the mean value has the form

$$\prod_{p \le x} \left(1 - \frac{1}{p-1} + \sum_{m \ge 1} \frac{g(p^m)}{p^m} \right) + o(1) \quad (x \to \infty), \tag{5.3}$$

where the product extended over all primes, converges. Hildebrand investigated the limit behavior assuming only the convergence of

$$\sum_p \frac{Reg(p) - 1}{p},$$

and he obtained that in this case (5.3) gives the right asymptotic for the partial sums as well. Using a partial summation then one has that if

$$\sum_p \frac{Reg(p)p^{-i\tau} - 1}{p}$$

converges, for a fixed real τ then

$$\pi(x)^{-1} \sum_{p \le x} g(p+1) = \frac{x^{i\tau}}{1 + i\tau} \prod_{p \le x} \left(1 - \frac{1}{p-1} + \sum_{m \ge 1} \frac{g(p^m)p^{-im\tau}}{p^m} \right) + o(1) \quad (x \to \infty).$$

This particular case corresponds to the famous result of Halász, therefore one could conjecture that if

$$\sum_p \frac{Reg(p)p^{-i\tau} - 1}{p}$$

diverges for all real τ then

$$\pi(x)^{-1} \sum_{p \le x} g(p+1) \to 0 \quad (x \to \infty).$$

Unfortunately Timofeev proved in [41] that if there is a primitive character χ_d modulo d such that

$$\sum_p \frac{1 - Re\chi_d(p)f(p)p^{-i\tau}}{p}$$

converges for some τ then

$$\begin{aligned}\frac{1}{\pi(x)}\sum_{n\le x} f(p+1) = &\frac{\mu(d)}{\varphi(d)}\frac{x^{i\tau}}{1+i\tau}\\ &\times \prod_{\substack{p\le x\\ p\nmid d}}\left(1+\sum_{r\ge 1}\frac{\chi_d(p^r)f(p^r)p^{-ri\tau} - \chi_d(p^{r-1})f(p^{r-1})p^{-(r-1)i\tau}}{\varphi(p^r)}\right)\\ &+ o(1)\end{aligned}$$

as $x \to \infty$, which is not necessarily 0 if χ_d is a real character.

Mean value estimates of this form remain true on the set $\mathcal{P}_k$+1. The main tool in proving the corresponding result will be the Fundamental Lemma of sieve theory. We will use it in the form given by Lemma 13. We proceed using the method of Chapter 4. inasmuch as we deduce the results from the analogous for $D\mathcal{P}+1$.

Let

$$M(x,f,D) := \sum_{Dp+1\le x} f(Dp+1).$$

Theorem 3. *Let $f(n)$ be a multiplicative function of modulus* 1. *Let furthermore d be positive integer. Suppose that there is a real τ such that the serie*

$$\sum_p \frac{|\chi(p)f(p)p^{i\tau}-1|^2}{p} \tag{5.4}$$

converge for some primitive character χ (mod d).

Then

$$\begin{aligned}\left(\pi\left(\frac{x-1}{D}\right)\right)^{-1} M(x,f,D) = &\frac{x^{i\tau}}{1+i\tau}\frac{\mu(d)}{\varphi(d)}\\ &\times\prod_{\substack{p\le x\\ p\nmid dD}}\left(1-\frac{1}{p-1}+\sum_{\alpha\ge 1}\frac{f(p^\alpha)p^{-i\alpha\tau}\chi(p^\alpha)}{p^\alpha}\right)\\ &+ o(1) \quad (x\to\infty)\end{aligned}$$

uniformly for all $D \le x^\varepsilon$ with $(d,D)=1$, where $0\le\varepsilon<1$.

Proof. First suppose that $\tau = 0$. We set $r = \log x$, and $x_D = \frac{x-1}{D}$. We use the notation $\log_2 x := \log\log x$, and $\log_3 x := \log\log_2 x$. First we note that using (4.15),(4.16) and the notation

$$f'(p^\alpha) = \begin{cases} f(p^\alpha) \text{ if } p \le r \text{ or } r < p \le x_D^{1-\vartheta_x},\ \alpha = 1 \\ 1 \text{ if } p > x_D^{1-\vartheta_x} \text{ or } r < p,\ \alpha \ge 2, \end{cases}$$

where ϑ_x is sequence tending slowly to 0 with $x \to \infty$ we have

$$\left| \sum_{Dp+1 \le x} f(Dp+1) - f'(Dp+1) \right| = \pi\left(x_D\right) o(1) \quad (x \to \infty).$$

Therefore we can identify f with f'. As before, it is easier to work with the truncated version of f, which is defined as

$$f_r(p^\alpha) = \begin{cases} f(p^\alpha) \text{ if } p \le r \\ \overline{\chi}(p^\alpha) \text{ if } r < p \le x_D^{1-\vartheta_x} \\ 1 \text{ otherwise }. \end{cases}$$

Next we give an alternative representation of $M(x, f_r, D)$. It can be written as follows

$$\sum_{Dp+1 \le x} f_r(Dp+1) = \sum_{\substack{m \le x+1 \\ P(m) \le r \\ (D,m)=1}} f(m) \sum_{\substack{p \le x_D \\ p \equiv l_D \pmod m \\ \left(\frac{Dp+1}{m}, \mathcal{P}(r)\right)=1}} \overline{\chi}\left(\frac{Dp+1}{m}\right), \tag{5.5}$$

where

$$\mathcal{P}(r) := \prod_{p \le r} p,$$

and l_D is the unique residue class satisfying

$$Dl_D \equiv -1 \pmod d.$$

Since $(D, md) = 1$, therefore there is a unique $l_d(mb-1) \pmod{md}$ residue class such that if

$$\frac{Dp+1}{m} = b + td$$

then

$$p \equiv l_d(mb-1) \pmod{md}.$$

Therefore the inner sum is

$$\sum_{\substack{Dp\le x\\ Dp\equiv -1 \pmod m\\ \left(\frac{Dp+1}{m},\mathcal{P}(r)\right)=1}} \overline{\chi}_d\left(\frac{Dp+1}{m}\right) = \sum_{\substack{b=1\\ (b,d)=1\\ (bm-1,d)=1}}^{d} \overline{\chi}_d(b) J_m(x,r,b),$$

where

$$J_m(x,r,b) = \\ \#\left\{p \le x_D \,:\, p \equiv l_d(mb-1) \pmod{md},\, \left(\frac{Dp+1}{m},\mathcal{P}(r)\right)=1\right\}.$$

Note that if $d<r$ then $J_m(x,r,b)$ is empty for all b with $(bm-1,d)\neq 1$. Furthermore, since

$$Dp+1 = m(b+td),$$

where $P(m)\le r$ and $(b+td,\mathcal{P}(r))=1$ therefore there is at most one m having the same parity as D for which $J_m(x,r,b)\neq 0$. We use the Fundamental Lemma of sieve theory (Lemma 13) to estimate $J_m(x,r,b)$.

In this Lemma the set $\{a_n:\ n=1\ldots N\}$ is now

$$\left\{\frac{Dp+1}{m} \le x/m :\ p\equiv l_d(bm-1) \pmod{dm}\right\},$$

for which the number of integers in this set is about

$$X = \frac{\pi\left(x_D\right)}{\varphi(dm)},$$

if dm is not to large. If $(\delta,d)>1$ or $(\delta,D)>1$ then since

$$Dp+1 = c\delta m \quad \text{and} \quad Dp+1 = mb+tdm,$$

therefore

$$\sum_{\substack{n=1\\ a_n\equiv 0 \pmod \delta}}^{N} 1 = 0,$$

such that we chose $\eta(\delta)=0$ in this case. Note that it was $(b,d)=1$.

In the other case there is a unique $l_\delta \pmod d$ with $\delta l_\delta \equiv b \pmod d$, therefore the above sum equals

$$\begin{aligned}&\#\{p\le x_D:\ Dp\equiv bm-1 \pmod{dm},\ Dp\equiv -1 \pmod{\delta m}\}\\ &=\#\{p\le x_D:\ Dp\equiv ml_\delta\delta-1 \pmod{m\delta d}\},\end{aligned}$$

which is about

$$\frac{\pi\left(x_D\right)}{\varphi(\delta dm)},$$

therefore we chose

$$\eta(\delta)=\frac{\varphi(dm)}{\varphi(\delta dm)}$$

in this case. Using Lemma 13 we obtain that

$$J_m(x,r,b)=\pi\left(x_D\right)\frac{1}{\varphi(dm)}\prod_{\substack{q\le r\\ p\nmid d\\ p\nmid D}}(1-\rho(q))(1+\theta H)+\mathcal{O}\left(\sum_{\substack{\delta\le\xi^3\\ P(\delta)\le r}}3^{\omega(\delta)}|\eta(x,\delta)|\right),$$

where θ, H is given in that lemma, and

$$\xi^3=\frac{x^{1/2}}{md\sqrt{D}\log^A x},$$

$$\eta(x,\delta)=\left(\sum_{\substack{p\le x_D\\ Dp\equiv bm-1 \pmod{dm}\\ Dp\equiv -1 \pmod{\delta m}}}1\right)-\rho(\delta)\frac{\pi\left(x_D\right)}{\varphi(dm)},$$

where A is an arbitrary positive fixed constant. We note that

$$|\eta(x,\delta)|\le\max_{(l,\delta md)=1}\left|\pi\left(x_D,\delta md,l\right))-\frac{\pi\left(x_D\right)}{\varphi(\delta dm)}\right|.$$

Since

$$S=\sum_{p\le r}\frac{\rho(p)}{1-\rho(p)}\log p\ll\log r,$$

therefore together with our assumption on r, $\max(\log r,S)\le\frac{1}{8}\log\xi$ holds, as well as

$$H=\mathcal{O}\left(\exp\left(-\frac{\log x}{\log r}\right)\right).$$

Furthermore (5.5) can be written as

$$\sum_{Dp+1\le x} f_r(Dp+1) = \pi\left(x_D\right) \sum_{\substack{m\le x^\kappa\\ P(m)\le r\\ (D,m)=1}}{}' f(m) \sum_{\substack{b=1\\ (b(bm-1),d)=1}}^{d} \overline{\chi}(b)\tfrac{1}{\varphi(dm)}$$

$$\times \prod_{\substack{p\le r\\ p\nmid dD}} \left(1-\tfrac{\varphi(dm)}{\varphi(pdm)}\right)\left(1+\mathcal{O}\left(\exp\left(-\tfrac{\log x}{\log r}\right)\right)\right)$$

$$+\mathcal{O}\left(\sum_{\substack{x^\kappa<m\le x\\ P(m)\le r}} \#\left\{Dp+1\le x:\ Dp+1\equiv 0 \pmod m\ \left(\tfrac{Dp+1}{m}, P(r)\right)=1\right\}\right)$$

$$+\mathcal{O}\left(\sum_{\delta\le \frac{\sqrt{x}}{\sqrt{D}\log^A x}} 3^{\omega(\delta)} \max_{(l,\delta)=1}\left|\pi\left(x_D,\delta,l\right)-\tfrac{\pi(x_D)}{\varphi(\delta)}\right|\right),$$

where κ is to be chosen such that $x^\kappa \le \frac{\sqrt{x}}{\sqrt{D}\log^A x}$ and $\sum'$ means that m and D is of opposite parity.

To estimate the first error term we note that all primes for which $Dp\equiv -1 \pmod m$ and $\left(\frac{Dp+1}{m},\mathcal{P}(r)\right)=1$, are counted only for one m in this term, therefore we have that it is not more than

$$\tfrac{1}{\kappa\log x}\sum_{P(m)\le r}\log m\#\left\{Dp+1\le x:\ Dp\equiv -1 \pmod m,\ \left(\tfrac{Dp+1}{m},\mathcal{P}(r)\right)=1\right\}$$

$$\le \tfrac{1}{\kappa\log x}\sum_{p^\alpha\le r}\pi(x_D,l_D,p^\alpha)\log p^r+\mathcal{O}\left(\tfrac{x-1}{\kappa D\log x}\sum_{\substack{p\ge\sqrt{r}\\ \alpha\ge 2}}\tfrac{\log p^r}{p^r}\right)$$

$$\ll \pi\left(x_D\right)\tfrac{\log r}{\log x}+\pi\left(x_D\right)o(1)\quad (x\to\infty).$$

The second error term is at most by means of Lemma 4

$$\pi\left(x_D\right)\log^{-A} x.$$

The main term can be extended over all integers, in the way that we show that the contribution of the sum over $m>x^\kappa$ is negligible i.e. it is

$$c\pi\left(x_D\right)\sum_{\substack{m>x^\kappa\\ P(m)\le r}}\frac{\varphi(d)}{\varphi(dm)}\prod_{\substack{p\le r\\ p\nmid dD}}\left(1-\frac{\varphi(dm)}{\varphi(pdm)}\right)$$

$$=c\pi\left(x_D\right)\prod_{p\le r}\left(1-\frac{1}{p}\right)\sum_{\substack{m>x^\kappa\\ P(m)\le r}}\frac{1}{\varphi(m)}\prod_{\substack{p|m\\ p|d}}\left(1-\frac{1}{p}\right)\prod_{p|d,\ p|m}\left(1-\frac{1}{p}\right)^{-1}$$

$$\times\prod_{p|dD,\ p\nmid m}\left(1-\frac{1}{p}\right)^{-1}\prod_{p\nmid d,\ p\nmid D,\ p\nmid m}\left(1-\frac{1}{p}\right)^{-1}\left(1-\frac{1}{\varphi(p)}\right).$$

Since

$$\begin{aligned}
\sum_{\substack{m>x^\kappa\\ P(m)\le r}} \frac{1}{\varphi(m)} &\ll \frac{1}{\kappa\log x} \sum_{\substack{m\\ P(m)\le r}} \frac{\log m}{\varphi(m)}\\
&= \frac{1}{\kappa\log x}\sum_{\substack{p\le r\\ \alpha\ge 1}} \log p^\alpha \sum_{\substack{m\\ P(m)\le r,\ p^\alpha || m}} \frac{1}{\varphi(m)}\\
&= \frac{1}{\kappa\log x}\sum_{\substack{p\le r\\ \alpha\ge 1}} \frac{\log p^\alpha}{\varphi(p^\alpha)} \sum_{\substack{m\\ P(m)\le r,\ (p,m)=1}} \frac{1}{\varphi(m)}\\
&\ll \frac{\log r}{\kappa\log x}\sum_{\substack{p\le r\\ \alpha\ge 1}} \frac{\log p^\alpha}{p^\alpha} \ll \frac{\log^2 r}{\kappa \log x},
\end{aligned}$$

therefore we see that

$$\begin{aligned}
\sum_{Dp+1\le x} f_r(Dp+1) = \pi\left(x_D\right) &\sideset{}{'}\sum_{\substack{m,\ P(m)\le r\\ (D,m)=1}} f(m) \sum_{\substack{b=1\\ (b(bm-1),d)=1}}^{d} \overline{\chi}(b)\frac{1}{\varphi(dm)}\\
&\times \prod_{\substack{p\le r\\ p\nmid dD}} \left(1-\frac{\varphi(dm)}{\varphi(pdm)}\right)\left(1+\mathcal{O}\left(\exp\left(-\frac{\log x}{\log r}\right)\right)\right)\\
&+\mathcal{O}(\varepsilon_2(x,r)),
\end{aligned} \tag{5.6}$$

where

$$\begin{aligned}
\varepsilon_2(x,r) &= \pi\left(x_D\right)\left(\prod_{p|dD,\ p\le r}\left(1-\frac{1}{p}\right)^{-1}\frac{\log r}{\kappa\log x}+o(1)\right)\\
&\ll \pi\left(x_D\right)\left(\frac{\log_2^2 x}{\log x}+o(1)\right)\\
&\ll \pi\left(x_D\right) o(1) \quad (x\to\infty).
\end{aligned}$$

Since by the inclusion-exclusion principle and by the orthogonality relation of the Dirichlet characters

$$\sum_{\substack{b=1\\ (b(bm-1),d)=1}}^{d} \overline{\chi}(b) = \sum_{(b,d)=1} \overline{\chi}(b)\sum_{\delta|d}\frac{\mu(\delta)}{\varphi(\delta)}\sum_{\chi \pmod \delta} \chi_\delta(bm),$$

therefore (5.6) can be written as

$$\pi\left(x_D\right) \sum_{\substack{b=1\\(b,d)=1}}^{d} \frac{\overline{\chi}(b)}{\varphi(d)} \sum_{\delta|d} \frac{\mu(\delta)}{\varphi(\delta)} \sum_{\chi \pmod{\delta}} \chi_\delta(b) \sideset{}{'}\sum_{\substack{m,\ P(m)\le r\\(D,m)=1}} \frac{f(m)\chi_\delta(m)\varphi(d)}{\varphi(dm)}$$
$$\times \prod_{\substack{p\le r\\p\nmid dD}} \left(1-\frac{\varphi(dm)}{\varphi(pdm)}\right)\left(1+\mathcal{O}\left(\exp\left(-\frac{\log x}{\log r}\right)\right)\right)$$
$$+\,o(1) \quad (x\to\infty).$$

Furthermore

$$\begin{aligned}(5.7)\quad &\sideset{}{'}\sum_{\substack{m,\ P(m)\le r\\(D,m)=1}} \frac{f(m)\chi_\delta(m)\varphi(d)}{\varphi(dm)} \prod_{\substack{p\le r\\p\nmid dD}} \left(1-\frac{\varphi(dm)}{\varphi(pdm)}\right) = \prod_{\substack{p\le r\\p\nmid 2dD}} \left(1-\frac{1}{\varphi(p)}\right)\\ &\times \sideset{}{'}\sum_{\substack{m,\ P(m)\le r\\(D,m)=1}} \frac{f(m)\chi_\delta(m)}{m} \prod_{p\nmid 2d,\ p|m} \left(1-\frac{1}{\varphi(p)}\right)^{-1}.\end{aligned}$$

Keeping in mind that m and D has opposite parity, we have that if D is odd then (5.7) can be written as

$$\prod_{\substack{p\le r\\p\nmid 2dD}} \left(1-\frac{1}{\varphi(p)}\right)\left(\sum_{\alpha\ge1}\frac{f(2^\alpha)\chi_\delta(2^\alpha)}{2^\alpha}\right)\times$$
$$\sum_{\substack{m,\ P(m)\le r\\(2D,m)=1}} \frac{f(m)\chi_\delta(m)}{m} \prod_{p\nmid 2d,\ p|m} \left(1-\frac{1}{\varphi(p)}\right)^{-1},$$

where we sum over a multiplicative function $h(m)$ for which

$$h(p^\alpha) = \begin{cases} \frac{f(p^\alpha)\chi_\delta(p^\alpha)}{p^\alpha}\left(1-\frac{1}{p-1}\right)^{-1} \text{ if } p\nmid 2Dd\\ \frac{f(p^\alpha)\chi_\delta(p^\alpha)}{p^\alpha} \text{ if } p|d,\ p\nmid 2D\\ 0 \text{ otherwise },\end{cases}$$

therefore in this case (5.7) is

$$\prod_{\substack{p\le r\\p\nmid Dd}} \left(1-\frac{1}{p-1}+\sum_{\alpha\ge1}\frac{f(p^\alpha)\chi_\delta(p^\alpha)}{p^\alpha}\right) \prod_{\substack{p\le r\\p\nmid 2D\\p|d}} \left(1+\sum_{\alpha\ge1}\frac{f(p^\alpha)\chi_\delta(p^\alpha)}{p^\alpha}\right),$$

One can see that if D is even then (5.7) has the same form as in the other case.

We have that

$$\begin{aligned}
&\pi\left(x_D\right)^{-1} M(x, f_r, D)\\
&= \sum_{\substack{b=1\\(b,d)=1}}^{d} \frac{\overline{\chi}(b)}{\varphi(d)} \sum_{\delta\mid d} \frac{\mu(\delta)}{\varphi(\delta)} \sum_{\chi \pmod{\delta}} \chi_\delta(b) \left(1 + \mathcal{O}\left(\exp\left(-\frac{\log x}{\log r}\right)\right)\right)\\
&\times \prod_{\substack{p\le r\\ p\nmid dD}} \left(1 - \frac{1}{p-1} + \sum_{\alpha\ge 1} \frac{f(p^\alpha)\chi_\delta(p^\alpha)}{p^\alpha}\right) \qquad (5.8)\\
&\times \prod_{\substack{p\le r\\ p\mid d,\ p\nmid 2D}} \left(1 + \sum_{\alpha\ge 1} \frac{f(p^\alpha)\chi_\delta(p^\alpha)}{p^\alpha}\right)\\
&+ o(1) \quad (x\to\infty).
\end{aligned}$$

The contribution arising from $\delta \neq d$ can be avoided as follows. Since the character induced by $\chi_\delta \cdot \overline{\chi}$ is not the principal character if $\chi_\delta \neq \chi$, therefore using Dirichlet's theorem in arithmetic progressions we have that

$$\sum_{z\le p\le r} \frac{|1-\chi_\delta\cdot\overline{\chi}(p)|^2}{p} \gg \frac{1}{\varphi(d)^2} \log\left(\frac{\log r}{\log z}\right) \gg \log\left(\frac{\log_2 x}{\log_3 x}\right),$$

if $z = \log_2 x$. Here we used that $\chi_\delta \cdot \overline{\chi}(p)$ is at most a $\varphi(d)$-th root of unity.

Since

$$|\chi_\delta(p)f(p) - 1|^2 \gg |1 - \overline{\chi}(p)\chi_\delta(p)|^2 - |1 - \overline{\chi}(p)f(p)|^2,$$

therefore

$$\begin{aligned}
\sum_{z\le p\le r} \frac{|1-\chi_\delta(p)f(p)|^2}{p} \gg& \sum_{z\le p\le r} \frac{|1-\chi_\delta(p)\overline{\chi}(p)|^2}{p}\\
&+ \mathcal{O}\left(\sum_{z\le p\le r} \frac{|1-\overline{\chi}(p)f(p)|^2}{p}\right)\\
\gg& \frac{1}{\varphi(d)^2} \log\left(\frac{\log_2 x}{\log_3 x}\right) + o(1) \quad (x\to\infty).
\end{aligned}$$

Furthermore using the inequality

$$|\log(1+a) - a| \le a^2,$$

as before we have that

$$\begin{aligned}\left|\prod_{\substack{p\le r\\ p\nmid dD}}\left(1-\frac{1}{p-1}+\sum_{\alpha\ge1}\frac{f(p^\alpha)\chi_\delta(p^\alpha)}{p^\alpha}\right)\right| &\ll \left|\exp\left(\sum_{p\le r}\frac{f(p)\chi_\delta(p)-1}{p}\right)\right|\\ &\ll \exp\left(\sum_{p\le r}\frac{Ref(p)\chi_\delta(p)-1}{p}\right)\\ &\ll \exp\left(-\sum_{z\le p\le r}\frac{|f(p)-\chi_\delta(p)|^2}{2p}\right)\\ &= o(1) \quad (x\to\infty).\end{aligned}$$

Putting it back into (5.8) we have that

$$\begin{aligned}\pi\left(x_D\right)^{-1}M(x,f_r,D) &= \frac{\mu(d)}{\varphi(d)}\\ &\quad\times\prod_{\substack{p\le r\\ p\nmid dD}}\left(1-\frac{1}{p-1}+\sum_{\alpha\ge1}\frac{f(p^\alpha)\chi(p^\alpha)}{p^\alpha}\right)\\ &\quad\times\prod_{\substack{p\le r\\ p\mid d,\ p\nmid 2D}}\left(1+\sum_{\alpha\ge1}\frac{f(p^\alpha)\chi(p^\alpha)}{p^\alpha}\right)\\ &\quad+o(1)\sum_{\substack{b=1\\(b,d)=1}}^{d}\frac{1}{\varphi(d)}\sum_{\delta\mid d}\frac{1}{\varphi(\delta)}\sum_{\chi\pmod\delta}1 \quad (x\to\infty),\end{aligned}\tag{5.9}$$

where the error term is not more than

$$o(1)\tau(d) \quad (x\to\infty).$$

Since $\chi(p^\alpha)=0$ for all $p|d$, therefore introducing the following notation

$$P(y):=\prod_{\substack{p\le y\\ p\nmid dD}}\left(1-\frac{1}{p-1}+\sum_{\alpha\ge1}\frac{f(p^\alpha)\chi(p^\alpha)}{p^\alpha}\right)$$

we have that

$$\pi\left(x_D\right)^{-1}M(x,f_r,D)=\frac{\mu(d)}{\varphi(d)}P(r)+o(1) \quad (x\to\infty).\tag{5.10}$$

Here we note that if (5.4) converges for $\tau = 0$ then $1 \ll |P(r)| \leq 1$. Now we can prove that

$$\frac{\mu(d)}{\varphi(d)} P\left(x_D\right)$$

is a good approximation to the sum $M(x, f, D)$.

Indeed

$$\begin{aligned}
&\left|\pi^{-1}(x_D) M(x, f, D) - \frac{\mu(d)}{\varphi(d)} P\left(x_D\right)\right| \\
&\qquad \leq \left|\pi^{-1}(x_D) M(x, f, D) - \pi^{-1}(x_D) M(x, f_r, D) \frac{P\left(x_D\right)}{P(r)}\right| \\
&\qquad\quad + \left|\frac{\mu(d)}{\varphi(d)} P\left(x_D\right) - \pi^{-1}(x_D) M(x, f_r, D) \frac{P\left(x_D\right)}{P(r)}\right|,
\end{aligned}$$

and using (5.10)

$$\left|\frac{\mu(d)}{\varphi(d)} P\left(x_D\right) - \pi^{-1}(x_D) M(x, f_r, D) \frac{P\left(x_D\right)}{P(r)}\right| = o(1) \quad (x \to \infty).$$

Therefore we have to show that

$$\left|\pi^{-1}(x_D) M(x, f, D) - \pi^{-1}(x_D) M(x, f_r, D) \frac{P\left(x_D^{\vartheta}\right)}{P(r)}\right| = o(1) \qquad (x \to \infty). \tag{5.11}$$

We note that if $d < r$ then

$$|f_r(p^\alpha)| = 1,$$

such that there is an additive function $g_r^*(p^\alpha) \in (-\pi, \pi]$ with

$$f_r^*(n) = f \cdot \overline{f}_r(n) = e^{i g_r^*(n)}.$$

Let

$$g_{r,1}^*(n) = \sum_{\substack{p^\alpha || n \\ p \leq x_D^{\varrho}}} g_r^*(p^\alpha),$$

and

$$g_{r,2}^*(n) = \sum_{\substack{p^\alpha || n \\ p > x_D^{\varrho}}} g_r^*(p^\alpha),$$

where $0<\varrho<1$ is fixed. We note that if

$$p \le r \quad \text{or} \quad p > x_D^{1-\vartheta_x} \quad \text{then} \quad g_r^*(p^\alpha)=0$$

for all natural number α. In order to prove (5.11) we use this functions to give a Turán-Kubilius type inequality. We have that

$$\sum_{Dp+1\le x}\left|g_r^*(Dp+1)-\sum_{\substack{q^m\le x_D\\ q\nmid D}}\frac{g_r^*(q^m)}{q^m}\right|^2 \ll \pi(x_D)\sum_{p^m\le x_D}\frac{|g_r^*(p^m)|^2}{p^m}. \tag{5.12}$$

To see this we split the summation on the left into three parts Σ_1, Σ_2, Σ_3 respectively, where

$$\Sigma_1=\sum_{Dp+1\le x}\left|g_{r,1}^*(Dp+1)-\sum_{\substack{q^m\le x_D^\varrho\\ q\nmid D}}\frac{g_r^*(q^m)}{q^m}\right|^2,$$

$$\Sigma_2=\sum_{Dp+1\le x}\left|g_{r,2}^*(Dp+1)\right|^2,$$

$$\Sigma_3=\sum_{Dp+1\le x}\left|\sum_{\substack{x_D^\vartheta<q^m\le x_D\\ q\nmid D}}\frac{g_r^*(q^m)}{q^m}\right|^2.$$

It is easy to see that the left hand side of (5.12) is at most $4(\Sigma_1+\Sigma_2+\Sigma_3)$. Σ_1 and Σ_2 can be estimated in the same manner as (4.19) and (4.20), and we get that they are not more than

$$c\pi\left(x_D\right)\sum_{\substack{p^m\le x_D\\ p\nmid D}}\frac{|g_r^*(p^m)|^2}{p^m}.$$

Further, using the Cauchy-Schwarz inequality we have that

$$\Sigma_3\ll\pi(x_D)\sum_{x_D^\vartheta<p^m\le x_D}\frac{1}{p^m}\sum_{x_D^\vartheta<p^m\le x_D}\frac{|g_r^*(p^m)|^2}{p^m}\ll\pi(x_D)\sum_{p^m\le x_D}\frac{|g_r^*(p^m)|^2}{p^m},$$

which proves (5.12).

Defining $A(x)$ to be the sum

$$\sum_{\substack{p^m \le x_D \\ p \nmid D}} \frac{g_r^*(p^m)}{p^m}$$

we have that (5.11) is

$$\begin{aligned}
&\pi(x_D)^{-1} \left| \sum_{Dp+1 \le x} f(Dp+1) - f_r(Dp+1) \frac{P(x_D)}{P(r)} \right| \\
&\quad \le \pi(x_D)^{-1} \sum_{Dp+1 \le x} \left| f_r^*(Dp+1) - \frac{P(x_D)}{P(r)} \right| \\
&\quad \le \pi(x_D)^{-1} \sum_{Dp+1 \le x} |f_r^*(Dp+1) - \exp[iA(x)]| \\
&\qquad + \left| \exp[iA(x)] - \frac{P(x_D)}{P(r)} \right| \\
&\quad = \Sigma_1' + \Sigma_2'.
\end{aligned}$$

Further using the Cauchy-Schwarz inequality again we have that

$$\begin{aligned}
\Sigma_1' &= \pi(x_D)^{-1} \sum_{Dp+1 \le x} |\exp[i(g_r^*(Dp+1) - A(x))] - 1| \\
&\le \pi(x_D)^{-1/2} \left(\sum_{Dp+1 \le x} |g_r^*(Dp+1) - A(x)|^2 \right)^{1/2}.
\end{aligned}$$

Using (5.12) we have that the right hand side of this last expression is at most

$$\left(\sum_{\substack{p^m \le x_D \\ p \nmid D}} \frac{|g_r^*(p^m)|^2}{p^m} \right)^{1/2}.$$

Furthermore since

$$|g_r^*(p^m)|^2 \ll |f_r^*(p^m) - 1|^2 = |f(p^m) - f_r(p^m)|^2,$$

therefore

$$\Sigma_1' \ll \left(\sum_{r \le p \le x} \frac{|\chi(p) f(p) - 1|^2}{p} \right)^{1/2} + \left(\sum_{r \le p \le x} \sum_{m \ge 2} \frac{1}{p^m} \right)^{1/2},$$

which according to the condition 5.4 clearly tends to zero as $r \to \infty$. We must consider Σ_2'. It can be written as

$$\left|1 - \prod_{\substack{r<p\le x_D\\ p\nmid D}} \left(1 - \frac{1}{p-1} + \sum_{m\ge 1} \frac{f(p^m)\chi(p^m)}{p^m}\right) \exp\left(-i \sum_{\substack{r<p\le x_D,\, p\nmid D\\ p^m\le x_D}} \frac{g_r^*(p^m)}{p^m}\right)\right|$$

$$= \left|1 - \exp\left[\sum_{\substack{r<p\le x_D\\ p\nmid D}} \frac{f(p)\chi(p)-1}{p} - i\frac{g_r^*(p)}{p} + \mathcal{O}\left(\frac{1}{p^2}\right)\right]\right|$$

$$= \left|1 - \exp\left[\mathcal{O}\left(\sum_{r<p\le x_D} \frac{|f(p)\chi(p)-1|^2}{p} + \sum_{r<p} \frac{1}{p^2}\right)\right]\right|,$$

which again tends to zero as $x \to \infty$. Finally we note that since $x^{1-\varepsilon} < x_D$ therefore we have

$$\begin{aligned}|P(x_D) - P(x)| &\ll \left|\prod_{x_D<p\le x} \left(1 + \frac{f(p)\chi(p)-1}{p} + \mathcal{O}\left(\frac{1}{p^2}\right)\right) - 1\right|, \\ &= \left|\exp\left(\sum_{x_D<p\le x} \frac{f(p)\chi(p)-1}{p} + \mathcal{O}\left(\frac{1}{p^2}\right)\right) - 1\right|,\end{aligned}$$

which tends to zero as $x \to \infty$ inasmuch as

$$\begin{aligned}\left|\sum_{x_D<p\le x} \frac{f(p)\chi(p)-1}{p}\right| &\ll \left(\sum_{x_D<p\le x} \frac{1}{p}\right)^{1/2} \left(\sum_{x_D<p\le x} \frac{|f(p)\chi(p)-1|^2}{p}\right)^{1/2} \\ &= o(1) \quad (x \to \infty).\end{aligned}$$

We have proved Theorem 3 in the case $\tau = 0$.

Now consider the case of an arbitrary τ. We proved that

$$\begin{aligned}\pi\left(x_D\right)^{-1} M(x, f(n)n^{-i\tau}, D) &= \frac{\mu(d)}{\varphi(d)} \\ &\times \prod_{\substack{p\le x\\ p\nmid dD}} \left(1 - \frac{1}{p-1} + \sum_{\alpha\ge 1} \frac{f(p^\alpha)p^{-i\alpha\tau}\chi(p^\alpha)}{p^\alpha}\right) \\ &+ o(1) =: \psi(x) + o(1)\end{aligned}$$

as $x \to \infty$.

Using a partial summation we get that

(5.13)
$$\sum_{Dp+1\le x} f(Dp+1) = x^{i\tau} \sum_{Dp+1\le x} f(Dp+1)(Dp+1)^{-i\tau} - i\tau \int_2^x \sum_{Dp+1\le u} f(Dp+1)(Dp+1)^{-i\tau} u^{i\tau-1} du.$$

Since if $D < x^{\varepsilon}$ then $D < x^{\gamma\varepsilon'}$ with some other $\varepsilon < \varepsilon' < 1$ and an appropriate $0 \le \gamma < 1$, therefore the estimation

$$\pi\left(\frac{u-1}{D}\right)^{-1} M(u, f(n)n^{-i\tau}, D) = \frac{\mu(d)}{\varphi(d)} \times \prod_{\substack{p\le u\\ p\nmid dD}} \left(1 - \frac{1}{p-1} + \sum_{\alpha\ge 1} \frac{f(p^{\alpha})p^{-i\alpha\tau}\chi(p^{\alpha})}{p^{\alpha}}\right) + o(1) \quad (x\to\infty)$$

remains valid in the range $x^{\gamma} < u < x$. Therefore we can estimate the integral on the right hand side of (5.13) in this range such that

(5.14)
$$\int_{x^{\gamma}}^{x} \sum_{Dp+1\le u} f(Dp+1)(Dp+1)^{-i\tau} u^{i\tau-1} du = \frac{\mu(d)}{\varphi(d)} \int_{x^{\gamma}}^{x} \pi(u_D)\psi(u)u^{i\tau-1} du + o(1) \int_{x^{\gamma}}^{x} \frac{1}{D\log u} du \quad (x\to\infty).$$

Now if $x^{\gamma} \le u \le x$ then using the Cauchy-Schwarz inequality

$$|\psi(x) - \psi(u)| \ll \sum_{u\le p\le x} \frac{|\overline{\chi}(p) - f(p)p^{-i\tau}|}{p} \ll \left(\sum_{u\le p\le x} \frac{1}{p}\right)^{1/2} \left(\sum_{u\le p\le x} \frac{|\overline{\chi}(p) - f(p)p^{-i\tau}|^2}{p}\right)^2 = o(1)$$

as $x \to \infty$, therefore the right hand side of (5.14) equals

(5.15)
$$\frac{\mu(d)}{\varphi(d)}\psi(x) \int_{x^{\gamma}}^{x} \pi(u_D)u^{i\tau-1} du + o(1) \int_{x^{\gamma}}^{x} \frac{1}{D\log u} du \quad (x\to\infty).$$

The error term here is at most

$$o(1)\,\frac{1}{D\log x}\int_{x^\gamma}^{x} du = o(\pi(x_D)) \quad (x\to\infty).$$

The integral appearing in the main term of (5.15) equals

$$\begin{aligned}\frac{1}{D}\int_{x^\gamma}^{x}\frac{u^{i\tau}}{\log(u/D)}du &= D^{i\tau}\int_{x^\gamma/D}^{x/D}\frac{u^{i\tau}}{\log u}du\\ &= \frac{D^{i\tau}}{1+i\tau}\left[\frac{u^{1+i\tau}}{\log u}\right]_{x^\gamma/D}^{x/D}+\mathcal{O}\left(\int_{x^\gamma/D}^{x/D}\frac{1}{\log^2 u}du\right)\\ &= \pi(x_D)\frac{x^{i\tau}}{1+i\tau}+o(\pi(x_D)) \quad (x\to\infty),\end{aligned}$$

therefore the left hand side of (5.14) equals

$$\pi(x_D)\frac{x^{i\tau}}{1+i\tau}\frac{\mu(d)}{\varphi(d)}\psi(x)+o(\pi(x_D)) \quad (x\to\infty).$$

Using the trivial bound

$$|M(u,f(n)n^{i\tau},D)|\le\pi(u_D),$$

we estimate the integral on the right hand side of (5.13) in the range $2\le u\le x^\gamma$ to be not more than

$$\mathcal{O}\left(\frac{1}{D}\int_{2D+1}^{x^\gamma}\frac{1}{\log(u/D)}du\right)\ll\int_{2}^{x^\gamma/D}\frac{1}{\log(u)}du=o(\pi(x_D)) \quad (x\to\infty).$$

Putting it all together we have that

$$\sum_{Dp+1\le x} f(Dp+1)=\pi(x_D)\frac{x^{i\tau}}{1+i\tau}\frac{\mu(d)}{\varphi(d)}\psi(x)+o(\pi(x_D)) \quad (x\to\infty),$$

as asserted.

□

As an application of Theorem 3 we are able to analyze the mean behavior of multiplicative functions on the set $\mathcal{P}_k+1$ in some cases.

Theorem 4. *Let $g(n)$ be a multiplicative function of modulus one, such that there is a primitive character χ (mod d) for some fixed d and a real τ such that*

$$\sum_p\frac{1-Re\chi(p)g(p)p^{-i\tau}}{p}$$

converges. Then

$$\pi_k(x)^{-1} \sum_{\substack{n\le x\\ \omega(n)=k}} g(n+1) = \frac{x^{i\tau}}{1+i\tau}\frac{\mu(d)}{\varphi(d)} \prod_{\substack{p\le x\\ p\nmid d}} \left(1-\frac{1}{p-1}+\sum_{\alpha\ge 1}\frac{f(p^\alpha)p^{-i\alpha\tau}\chi(p^\alpha)}{p^\alpha}\right) + o(1) \qquad (x\to\infty)$$

uniformly for all k having property $A(\varepsilon, x)$.

Proof. Using Lemma 8 and the notations of Lemma 1 we have that

$$\begin{aligned}\sum_{\substack{n\le x\\ \omega(n)=k}} g(n+1) &= \sum_{\pi_{k-1}\in S_k} \sum_{p^*_{\pi_{k-1}}\le p\le \frac{x}{\pi_{k-1}}} g(\pi_{k-1}p+1) + o(\pi_k(x)) \\ &= \sum_{\pi_{k-1}\in S_k} M(g,x,\pi_{k-1}) - \sum_{p\le p^*_{\pi_{k-1}}} g(\pi_{k-1}p+1) + o(\pi_k(x)) \\ &\qquad (x\to\infty).\end{aligned} \tag{5.16}$$

Let

$$\begin{aligned}\psi(x,D) :=& \frac{x^{i\tau}}{1+i\tau}\frac{\mu(d)}{\varphi(d)} \\ &\times \prod_{\substack{p\le x\\ p\nmid dD}} \left(1-\frac{1}{p-1}+\sum_{\alpha\ge 1}\frac{f(p^\alpha)p^{-i\alpha\tau}\chi(p^\alpha)}{p^\alpha}\right)\end{aligned}$$

Note that using Lemma 8 we have $M_x \le P(\pi_{k-1})$, therefore in our case π_{k-1} and d are coprimes for large x. By meanings of the remark after Lemma 9 we have that the second sum on the most right hand of (5.16) is $o(\pi_k(x))$. To estimate the first sum here we can apply Theorem 3 since for $\pi_{k-1} \in S_k$ we have $\pi_{k-1} \le x^{1/A_x}$, and we get that

$$\sum_{\substack{n\le x\\ \omega(n)=k}} g(n+1) = \sum_{\pi_{k-1}\in S_k} \psi(x,\pi_{k-1})\pi(x_{\pi_{k-1}}) + o(\pi_k(x)) \quad (x\to\infty).$$

Defining $K(x, D)$ by the identity

$$\psi(x,1) = \psi(x,D)K(x,D),$$

such that

$$K(x,D) = \prod_{\substack{p\le x\\ p\mid D}} \left(1-\frac{1}{p-1}+\sum_{\alpha\ge 1}\frac{f(p^\alpha)p^{-i\alpha\tau}\chi(p^\alpha)}{p^\alpha}\right),$$

holds, we have that the left hand side of (5.16) equals

$$\begin{aligned}
&\psi(x,1) \sum_{\pi_{k-1}\in S_k} \pi\left(x_{\pi_{k-1}}\right) \\
&+ \sum_{\pi_{k-1}\in S_k} \pi\left(x_{\pi_{k-1}}\right) \psi\left(x,\pi_{k-1}\right)\left[1-K(x,\pi_{k-1})\right] \\
&+ o(\pi_k(x)) \quad (x\to\infty).
\end{aligned}$$

Since using Lemma 8 we have $M_x \le p(\pi_{k-1})$ and since

$$K(x,\pi_{k-1}) = \exp\left[\sum_{\substack{p\le x\\ p\mid \pi_{k-1}}} \frac{f(p^\alpha)\chi(p^\alpha)p^{i\tau}}{p} + \mathcal{O}\left(\sum_{\substack{p\le x\\ p\mid \pi_{k-1}}} \frac{1}{p^2}\right)\right]$$

therefore the right hand side of (5.16) equals

$$\begin{aligned}
\psi(x,1) \sum_{\pi_{k-1}\in S_k} \pi\left(x_{\pi_{k-1}}\right) + o(1) \sum_{\pi_{k-1}\in S_k} \pi\left(x_{\pi_{k-1}}\right)& \\
+ o(\pi_k(x)) \quad (x\to\infty)&,
\end{aligned}$$

and using (4.23) the assertion follows.

□

5.2 Outlook, open questions

Mean value estimations of multiplicative functions with modulus at most one, play a central role in the theory of numbers, and has been investigated intensively. Most of the problems can be simplified to the investigation of the behavior of the multiplicative function g on prime values. For example asymptotic estimations for sums like

$$\sum_{n\le x} g(n)\overline{g(n+1)}, \tag{5.17}$$

have been made in [19], [20] for the case if

$$\sum_p \frac{|g(p)p^{i\tau}-1|^2}{p},$$

converges for some real τ and in [38] for the case $g(p)$ is always near in some sense to a fixed complex number. In both cases the convolution technique described in [28], [19] - and what we used in Chapter 4 - helps. Although there is a hope to obtain some new results with the method described in this Chapter if g is regularized by a Dirichlet character, the most interesting case concerning the Liouville function remains open. This latter having been conjectured to be deeply connected with the twin prime conjecture. The best known result in this direction is the following

$$-(1+o(1))\frac{1}{3} < \frac{1}{x}\sum_{n\le x}\lambda(n)\lambda(n+1) < 1 - \frac{1}{(\log x)^{7+\varepsilon}},$$

for $x > x(\varepsilon)$. See [13]. Our knowledge concerning (5.17) is not strong enough to be able to give asymptotic estimations as the following problem of Kátai and Subbarao shows (See [30], [31]). Suppose that the range of g is the multiplicative group of the k-th roots of unity with some k, and g is completely multiplicative. Is it true that $g(n)\overline{g(n+1)}$ takes all values in this group infinitely often? Currently this question is answered for $k \le 3$ in [30] and for $k \le 5$ in [33]. The answer to the above mentioned question is of course not new in the case of the Liouville function. In this sense the mean value problem

$$\sum_{p\le x} g(p+1)$$

seems to lie deeper, since we don't even know if both $\lambda(p+1) = 1$, and $\lambda(q+1) = -1$ holds infinitely often.

A future path in the investigation concerning the topic of this work could be to extend the results - based on Lemma 7- to a larger domain of k, for example for $1 \le k \le A\log\log x$.

Chapter 6

The Erdős-Kac Theorem

The central limit theorem in probability theory states that the sum of a large number of independent random variables each with finite mean and variance will be approximately normally distributed [1]. In Chapter 3 we have seen that strongly additive functions are "almost independent" random variables, therefore it is reasonably to expect that under some circumstances the centralized, normalized himself satisfies the central limit theorem. Erdős-Kac type theorems are well known in number theory and are analog of the central limit theorem for strongly additive functions. For the development of this topic see for example [18], [10] and [11], [2] as well as [26] and [9]. In what follows $G(z)$ will stand for the Gaussian distribution function. In the case of shifted primes this theorem is as follows

Theorem (Barban). *Let $f(m)$ be a real strongly additive function satisfying*

$$\mu_x = \max_{p \leq x} |f(p)| B(x)^{-1} \to 0 \quad (x \to \infty),$$

where

$$B(x) = (\sum_{p \leq x} \frac{f^2(p)}{p-1})^{1/2}.$$

Then we have with

$$A(x) = \sum_{p \leq x} \frac{f(p)}{p-1}$$

that

$$\nu_x(p \leq x : f(p+1) - A(x) \leq zB(x)) \Rightarrow G(z)$$

as $x \to \infty$.

The corresponding result for subsets of $\mathcal{P}_k(x)$ appeared also in [2] in a slightly different form

Theorem (Barban)**.** *Let $f(m)$ be a real strongly additive function satisfying*

$$\mu_x = \max_{p \le x} |f(p)| B(x)^{-1} \to 0 \quad (x \to \infty),$$

where

$$B(x) = (\sum_{p \le x} \frac{f^2(p)}{p-1})^{1/2}.$$

Then we have with

$$A(x) = \sum_{p \le x} \frac{f(p)}{p-1}$$

that

$$\nu_x(p_1 \le x^{\alpha_1}, p_2 \le x^{\alpha_2}, \ldots, p_k \le x^{\alpha_k} : \frac{f(p_1 p_2 \cdots p_k + 1) - A(x)}{B(x)} \le z) \Rightarrow G(z) \qquad (x \to \infty),$$

where $\alpha_1 + \alpha_2 + \cdots + \alpha_k = 1$.

In this Chapter we prove the following theorem

Theorem 5. *Let $f(m)$ be a real additive function satisfying*

$$\frac{1}{B^2(x)} \sum_{\substack{p \le x \\ |f(p)| > \varepsilon B(x)}} \frac{f^2(p)}{p-1} \to 0 \quad (x \to \infty), \tag{6.1}$$

for all fixed $\varepsilon > 0$ where

$$B(x) = (\sum_{p \le x} \frac{f^2(p)}{p-1})^{1/2}.$$

Then we have with

$$A(x) = \sum_{p \le x} \frac{f(p)}{p-1}$$

that

$$\nu_x(n \in \mathcal{P}_k(x) : \frac{f(n+1) - A(x)}{B(x)} \le z) \Rightarrow G(z) \quad (x \to \infty) \tag{6.2}$$

uniformly for all k having property $A(\varepsilon, x)$.

With the notations of the above theorem, we say that $f(n)$ belongs to the class H if there exists a function $r = r(x)$ such that

$$\frac{\log r}{\log x} \to 0, \quad \frac{B(r)}{B(x)} \to 1, \quad B(x) \to \infty$$

as $x \to \infty$. This class of functions was introduced by Kubilius. We proceed similarly as in the previous chapters. First we prove the following

Theorem 6. *Let $f(m)$ be an additive function of class H. Let*

$$B_D(x) = (\sum_{\substack{p \le x \\ p \nmid D}} \frac{f^2(p)}{p-1})^{1/2},$$

and let $\delta(x)$ be a sequence tending to zero arbitrary slowly. Furthermore let $0 < \sigma \le 1$ and let

$$(T_\delta(x) =)\ T(x) = \{D \le x^{1-\sigma} : B_D(x) = B(x)(1 + \mathcal{O}(\delta(x)))\}.$$

Then we have with

$$A_D(x) = \sum_{\substack{p \le x \\ p \nmid D}} \frac{f(p)}{p-1}$$

that

$$\nu_x(n \in K_D(x) : \frac{f(n) - A_D(x)}{B_D(x)} \le z) \Rightarrow G(z) \quad (x \to \infty)$$

uniformly for all positive integer $D \in T(x)$, if and only if for each fixed $\varepsilon > 0$

$$\frac{1}{B_D^2(x)} \sum_{\substack{p \le x \\ p \nmid D \\ |f(p)| > \varepsilon B_D(x)}} \frac{f^2(p)}{p} \to 0 \quad (x \to \infty) \tag{6.3}$$

uniformly for all $D \in T(x)$.

For the proof of Theorem 5 we need the following

Lemma 25. *For a sequence of distribution functions $F_n(x)$ the relations*

$$F_n(\alpha_n x + a_n) \Rightarrow F(x),$$
$$F_n(\beta_n x + b_n) \Rightarrow F(x),$$

as $n \to \infty$, where $\alpha_n, \beta_n > 0$, a_n, b_n are real numbers and $F(x)$ is a proper distribution function, are satisfied simultaneously if and only if

$$\frac{\alpha_n}{\beta_n} \to 1, \quad \frac{a_n - b_n}{\alpha_n} \to 0$$

as $n \to \infty$.

For a proof see for example Theorem 2. of §10 [15]. Using Theorem 6. we may proceed in the following way with the

Proof of Theorem 5. First we note that it is enough to prove the theorem only for strongly additive functions. Indeed, if we define $h(p^\alpha) = f(p)$ for all primes p, and all positive integers α, then by (2.2) f and h have the same limit distribution.

As in the proof of Theorem 1. we have that the characteristic function of the frequencies on the right hand side of (6.2) equals

$$\frac{1}{B_k(x)} e^{-it\frac{A(x)}{B(x)}} \sum_{\pi_{k-1}\in S_x} \sum_{p^*(\pi_{k-1})<p\leq\frac{x}{\pi_{k-1}}} e^{it\frac{f(\pi_{k-1}p+1)}{B(x)}}.$$

We write it in the following way

(6.4)
$$\frac{1}{B_k(x)} \sum_{\pi_{k-1}\in S_x} \sum_{p\leq\frac{x}{\pi_{k-1}}} e^{it\frac{f(\pi_{k-1}p+1)-A_{\pi_{k-1}}(x)+A_{\pi_{k-1}}(x)-A(x)}{B(x)}} + \mathcal{O}(\frac{1}{B_k(x)} \sum_{\pi_{k-1}\in S_x} \pi(p^*)).$$

The second term like in the proof of Theorem 1. is $o(B_k(x))$. We estimate the main term. Since $M_x \leq p(\pi_{k-1})$ therefore an application of the Cauchy-Schwarz inequality shows that

$$\begin{aligned} |A(x) - A_{\pi_{k-1}}(x)| &\leq \sum_{p|\pi_{k-1}} \frac{|f(p)|}{p} \\ &\leq B(x)(\sum_{M_x\leq p\leq M_x+k} \frac{1}{p})^{1/2} \\ &= o(B(x)) \end{aligned}$$

as $x \to \infty$. Substituting into (6.4) we have that it equals

$$\frac{1}{B_k(x)} \sum_{\pi_{k-1}\in S_x} \sum_{p\leq\frac{x}{\pi_{k-1}}} e^{it\frac{f(\pi_{k-1}p+1)-A_{\pi_{k-1}}(x)}{B(x)}} + o(1) \quad (x\to\infty). \tag{6.5}$$

Since

$$\begin{aligned} \sum_{\substack{p|\pi_{k-1}\\ |f(p)|\leq\varepsilon B(x)}} \frac{f^2(p)}{p} &\ll \varepsilon^2 B^2(x) \sum_{p|\pi_{k-1}} \frac{1}{p} \\ &\ll \varepsilon^2 B^2(x) \sum_{M_x\leq p\leq M_x+k} \frac{1}{p} \\ &= o(B^2(x)) \quad (x\to\infty), \end{aligned}$$

for each fixed $\varepsilon > 0$, therefore according to the condition (6.1)

$$B^2(x) - B^2_{\pi_{k-1}}(x) = \sum_{\substack{p|\pi_{k-1} \\ |f(p)| \le \varepsilon B(x)}} \frac{f^2(p)}{p} + \sum_{\substack{p|\pi_{k-1} \\ |f(p)| > \varepsilon B(x)}} \frac{f^2(p)}{p}$$
$$= \gamma(x) B^2(x)$$

for a typical $\pi_{k-1} \in S_x$, where $\gamma(x)$ is tending appropriate slowly to zero as $x \to \infty$. It follows that for each fixed $\varepsilon > 0$

$$\frac{1}{B^2_{\pi_{k-1}}(x)} \sum_{\substack{p \le x \\ p \nmid \pi_{k-1} \\ |f(p)| > \varepsilon B_{\pi_{k-1}}(x)}} \frac{f^2(p)}{p} \ll \frac{1}{B^2(x)} \sum_{\substack{p \le x \\ |f(p)| > \varepsilon (1+o(1)) B(x)}} \frac{f^2(p)}{p}$$
$$= o(1) \quad (x \to \infty),$$

consequently the conditions of Theorem 6. are fullfiled, and it is applicable for

$$T_\gamma(x) = \{\pi_{k-1} \in S_x\}.$$

Here we note that (6.1) implies that for all $r \le x^{1/2}$ and for all fixed $\varepsilon > 0$

$$\sum_{r \le p \le x} \frac{f^2(p)}{p} = \sum_{\substack{r \le p \le x \\ |f(p)| \le \varepsilon B(x)}} \frac{f^2(p)}{p} + \sum_{\substack{r \le p \le x \\ |f(p)| > \varepsilon B(x)}} \frac{f^2(p)}{p}$$
$$\ll \varepsilon^2 B^2(x) \log \frac{\log x}{\log r} + o(B^2(x)) \quad (x \to \infty),$$

therefore $f(n)$ is of class H. We obtained that

$$\nu_x\left(n \in K_{\pi_{k-1}}(x) : \frac{f(n+1) - A_{\pi_{k-1}}(x)}{B_{\pi_{k-1}}(x)} \le z\right) \Rightarrow G(z)$$
$$(x \to \infty),$$

uniformly for all $\pi_{k-1} \in S_x$. Applying Lemma 25. we deduce that

$$\nu_x\left(n \in K_{\pi_{k-1}}(x) : \frac{f(n+1) - A_{\pi_{k-1}}(x)}{B(x)} \le z\right) \Rightarrow G(z)$$
$$(x \to \infty),$$

uniformly for all $\pi_{k-1} \in S_x$ as well. Since the characteristic function of G is

$$\psi_G(t) = e^{-it^2/2},$$

therefore by meanings of Theorem 6. keeping in mind that $\pi_{k-1} \le x^{1/2}$ say, we have that the inner sum in the main term of (6.5) equals

$$\pi(\frac{x}{\pi_{k-1}})\psi(t) + o(\pi(\frac{x}{\pi_{k-1}})) \quad (x \to \infty).$$

Argumenting like in (4.23) the proof is finished. □

A direct consequence of Theorem 5 is the following

Corollary 1.

$$\nu_x(n \le x, \omega(n) = k : \frac{\omega(n+1) - \log\log x}{\sqrt{\log\log x}} \le z) \Rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-w^2/2} dw \quad (x \to \infty)$$

uniformly for all k having property $A(\varepsilon, x)$.

It remains to prove Theorem 6. Before we proceed we introduce a simple result which is useful in many cases. For a proof see for example Lemma 4.1 in [26]

Lemma 26. *Let $g_n(m), h_n(m)$ $(n = 1, 2, \ldots; m = 1, 2, \ldots, n)$ be two sequences of real numbers, which satisfies*

$$\nu_n(|g_n(m) - h_n(m)| \ge \varepsilon) \to 0 \quad (n \to \infty)$$

for all fixed $\varepsilon > 0$. We have that

$$\nu_n(g_n(m) \le z)$$

converge weakly as $n \to \infty$ if and only if

$$\nu_n(h_n(m) \le z)$$

converge weakly as $n \to \infty$.

Proof of Theorem 6. First we note that it is enough to prove the theorem only for strongly additive functions. Indeed, if we define $h(p^\alpha) = f(p)$ for all primes p, and all positive integers α and $t(n) = h(n) - f(n)$ for all positive integers n, then

$$\begin{aligned}\nu_x(n \in K_D(x) : |t(n)| > \kappa) \ll &\nu_x(n \in K_D(x) : |t(n)| > \kappa,\ p^2|n \Rightarrow p \le \delta_1) \\ &+ \nu_x(n \in K_D(x) : |t(n)| > \kappa,\ \exists p^2|n \Rightarrow p > \delta_1),\end{aligned}$$

for all $\delta_1 > 0$. By (4.16) the second therm is at most ϵ_{δ_1}, where $\epsilon_{\delta_1} \to 0$ as $\delta_1 \to \infty$, whilst the first term does not exceed

$$\pi^{-1}(\frac{x-1}{D}) \sum_{\substack{Dp+1 \le x \\ \exists q^2 | Dp+1 \Rightarrow q \le \delta_1 \\ |t(Dp+1)| > \kappa}} 1 \ll \sum_{p \le \delta_1} \sum_{\substack{2 \le \alpha \le \frac{\log \delta_2}{\log p} \\ |t(p^\alpha)| > \kappa}} \frac{1}{p^\alpha} + \epsilon_{\delta_2}$$

for all $\delta_2 > 2$, where

$$\begin{aligned}
\epsilon_{\delta_2} \ll & \pi^{-1}(\frac{x-1}{D}) \sum_{\substack{Dp+1 \le x \\ \exists q^\beta | Dp+1 \\ \beta > \frac{\log \delta_2}{\log q}}} 1 \\
\ll & \pi^{-1}(\frac{x-1}{D}) \sum_{\substack{Dp+1 \le x \\ \exists q^\beta | Dp+1 \\ \frac{\log \frac{x-1}{D}}{2 \log q} > \beta > \frac{\log \delta_2}{\log q}}} 1 + \pi^{-1}(\frac{x-1}{D}) \sum_{\substack{Dp+1 \le x \\ \exists q^2 | Dp+1 \\ q > \sqrt{\frac{x-1}{D}}}} 1 \\
= & o(1) \quad (\delta_2 \to \infty).
\end{aligned}$$

We obtained that

$$\overline{\lim}_{\kappa \to \infty} \overline{\lim}_{x \to \infty} \pi^{-1}(\frac{x-1}{D}) \nu_x(n \in K_D(x) : |t(n)| > \kappa) = 0.$$

It follows that

$$\pi^{-1}(\frac{x-1}{D}) \nu_x(n \in K_D(x) : |t(n)| > \epsilon B_1(x)) \to 0 \quad (x \to \infty)$$

for each $\epsilon > 0$, therefore using Lemma 26. f and h have the same limit distribution.

Set $r = x^\delta$, for some fixed $0 < \delta < \sigma \varrho$, where $0 < \varrho < 1/4$. Define strongly additive functions f_0, f_1, f_2 as follows.

$$f_0(p) = \begin{cases} f(p) & \text{if } p \le r \\ 0 & \text{otherwise,} \end{cases}$$

$$f_1(p) = \begin{cases} f(p) & \text{if } r < p \le \left(\frac{x-1}{D}\right)^\varrho \\ 0 & \text{otherwise,} \end{cases}$$

and let

$$f_2(p) = \begin{cases} f(p) & \text{if } \left(\frac{x-1}{D}\right)^\varrho < p \le \left(\frac{x-1}{D}\right)^{1-\vartheta_x} \\ 0 & \text{otherwise,} \end{cases}$$

where ϑ_x is tending to zero slowly as $x \to \infty$.

According to (4.15) we have that

$$\nu_x(n \in K_D(x) : f(n) - A_D(x) \leq zB_D(x)) \tag{6.6}$$

and

$$\nu_x(n \in K_D(x) : f_0(n) + f_1(n) + f_2(n) - A_D(x) \leq zB_D(x))$$

differs only by an $o(1)$ amount as $x \to \infty$. Since $f(n)$ is of class H and $D \in T(x)$, after an application of the Cauchy-Schwarz inequality we obtain

$$\begin{aligned}(\sum_{\substack{(\frac{x-1}{D})^\varrho \leq p \leq x \\ p \nmid D}} \frac{f(p)}{p})^2 &\ll \sum_{(\frac{x-1}{D})^\varrho \leq p \leq x} \frac{f^2(p)}{p} \\ &\ll o(B_D^2(x)) \quad (x \to \infty).\end{aligned}$$

It follows using Chebyshev's inequality that

$$\nu_x(n \in K_D(x) : |A_D(x) - A_D((\frac{x-1}{D})^\varrho)| > \varepsilon B_D(x))$$

does not exceed

$$\frac{1}{\varepsilon^2 B_D^2(x)}(\sum_{\substack{(\frac{x-1}{D})^\varrho \leq p \leq x \\ p \nmid D}} \frac{f(p)}{p})^2 = o(1) \quad (x \to \infty).$$

Applying Lemma 26. we deduce that (6.6) coincide with

$$\nu_x(n \in K_D(x) : f(n) - A_D((\frac{x-1}{D})^\varrho) \leq zB_D(x)) + o(1) \qquad (x \to \infty).$$

Exactly in the same way as in the proof of (4.19) we have that

$$\sum_{Dp+1 \leq x} |f_1(Dp+1) + A_D(r) - A_D((\frac{x-1}{D})^\varrho)|^2 \ll \pi(\frac{x-1}{D}) \sum_{\substack{r \leq p \leq (\frac{x-1}{D})^\varrho \\ p \nmid D}} \frac{f^2(p)}{p-1},$$

and like (4.20) we have

$$\sum_{Dp+1 \leq x} f_2^2(Dp+1) \ll \pi(\frac{x-1}{D}) \sum_{\substack{(\frac{x-1}{D})^\varrho < p \leq (\frac{x-1}{D})^{1-\vartheta_x} \\ p \nmid D}} \frac{f^2(p)}{p-1}.$$

Let

$$h(n) = f_1(n) + f_2(n) + A_D(r) - A_D((\frac{x-1}{D})^{\varrho})).$$

Using the above estimations after an application of Chebyshev's inequality for probabilities we have that for every $\varepsilon > 0$

$$\begin{aligned}\nu_x(n \in K_D(x) : |h(n)| > \varepsilon B_D(x)) \leq & \frac{1}{\varepsilon^2 \pi(\frac{x-1}{D}) B_D^2(x)} \sum_{n \in K_D(x)} |h(n)|^2 \\ \ll & \frac{1}{\varepsilon^2 B_D^2(x)} \sum_{\substack{r \leq p \leq (\frac{x-1}{D})^{1-\vartheta_x} \\ p \nmid D}} \frac{f^2(p)}{p-1}.\end{aligned}$$

Since by the conditions

$$B_D((\frac{x-1}{D})^{1-\vartheta_x}) - B_D(r) = o(B_D(x)) \quad (x \to \infty),$$

therefore in the view of Lemma 26. we have that

$$\nu_x(n \in K_D(x) : f(n) - A_D((\frac{x-1}{D})^{\varrho})) \leq z B_D(x))$$

equals with

$$\nu_x(n \in K_D(x) : f_0(n) - A_D(r) \leq z B_D(x)) + o(1) \quad (x \to \infty).$$

Further by Lemma 17. it equals with

$$\mathbb{P}(\sum_{\substack{p \leq r \\ p \nmid D}} X_p - A_D(r) \leq z B_D(x)) + \mathcal{O}(\exp(-\frac{\log x}{\log r}))$$

if δ is small enough and where the independent random variables X_p for all p are defined as follows:

$$X_p = \begin{cases} f(p) & \text{with probability } \frac{1}{p-1} \\ 0 & \text{with probability} 1 - \frac{1}{p-1}. \end{cases}$$

Since again by the conditions

$$\begin{aligned}\mathbb{P}(|\sum_{\substack{r \leq p \leq x \\ p \nmid D}} X_p + A_D(r) - A_D(x)| > \varepsilon B_D(x)) \leq & \frac{1}{\varepsilon^2 B_D^2(x)} \sum_{\substack{r \leq p \leq x \\ p \nmid D}} \frac{f^2(p)}{p-1} \\ = & o(1) \quad (x \to \infty),\end{aligned}$$

therefore we have that

$$\nu_x(n \in K_D(x) : f(n) - A_D(x) \le zB_D(x)) \Rightarrow F(z)$$

if and only if

$$\mathbb{P}(\sum_{\substack{p\le x\\ p\nmid D}} X_p - A_D(x) \le zB_D(x)) \Rightarrow F(z).$$

We follow the arguments of the proof [9] Theorem 12.1. The characteristic function of the distribution function on the left hand side above equals

$$\psi(t,x) = \exp(-it\frac{A_D(x)}{B_D(x)}) \prod_{\substack{p\le x\\ p\nmid D}} (1 + \frac{e^{it\frac{f(p)}{B_D(x)}} - 1}{p-1}).$$

By setting

$$K_D(x,u) = \frac{1}{B_D^2(x)} \sum_{\substack{p\le x\\ p\nmid D\\ f(p)\le uB_D(x)}} \frac{f^2(p)}{\varphi(p)}$$

we have that

$$\begin{aligned}
\log\psi(x,t) &= \sum_{\substack{p\le x\\ p\nmid D}} \varphi^{-1}(p)\{e^{it\frac{f(p)}{B_D(x)}} - 1 - it\frac{f(p)}{B_D(x)}\} + \mathcal{O}(\sum_{\substack{p\le x\\ p\nmid D}} \frac{f^2(p)}{\varphi^2(p)B_D^2(x)})\\
&= \int_{-\infty}^{\infty} \{e^{itu} - 1 - itu\}u^{-2}dK_D(x,u) + o(1) \quad (x\to\infty).
\end{aligned}$$

In this last step we used that for all fixed q we have

$$\begin{aligned}
\sum_{\substack{p\le x\\ p\nmid D}} \frac{f^2(p)}{\varphi^2(p)B_D^2(x)} &= \sum_{\substack{p\le q\\ p\nmid D}} \frac{f^2(p)}{\varphi^2(p)B_D^2(x)} + \sum_{\substack{p\ge q\\ p\nmid D}} \frac{f^2(p)}{\varphi^2(p)B_D^2(x)}\\
&\ll \frac{B_D^2(q)}{B_D^2(x)} + \frac{1}{q}\frac{1}{B_D^2(x)} \sum_{\substack{p\le x\\ p\nmid D}} \frac{f^2(p)}{\varphi(p)}\\
&\ll \frac{1}{q} + o(1) \quad (x\to\infty).
\end{aligned}$$

Since

$$e^{-it^2/2} = \int_{\mathbb{R}} \{e^{itu} - 1 - itu\}u^{-2}dK(u),$$

where

$$K(u) = \begin{cases} 1 & \text{if } u \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

therefore by the uniqueness of the representation by Kolmogorov's formula (see Theorem 3. of §19 in [15]) we have that

$$\int_{\mathbb{R}} \{e^{itu} - 1 - itu\} u^{-2} dK_D(x,t) \to \int_{\mathbb{R}} \{e^{itu} - 1 - itu\} u^{-2} dK(t) \qquad (x \to \infty),$$

if and only if

$$K_D(x,u) \Rightarrow K(u) \quad (x \to \infty).$$

It is easy to see that this last relation is equivalent to (6.3), and the proof is ready. $\square$

Bibliography

[1] http://en.wikipedia.org/wiki/Central_limit_theorem.

[2] M. B. Barban, *The 'Large Sieve' method and its applications in the theory of numbers*, Russ. Math. Surv. **21** (1966), 49–103.

[3] P. Billingsley, *The probability theory of additive arithmetical functions*, The Ann. of Prob. **2** (1974), no. 5, 749–791.

[4] ———, *Probability and measure*, 3 ed., New York, Toronto, London: John Wiley and Sons, 1995.

[5] H. Daboussi and K-H. Indlekofer, *Two elementary proofs of Halász's theorem*, Math. Zeitschrift **209** (1992), 43–52.

[6] H. Davenport, *Multiplicative number theory* , 3 ed., Springer-Verlag, New York, 2000.

[7] H. Delange, *Sur les fonctions arithmetiques multiplicatives*, Ann. Scient. EC. Norm. Sup. **78** (1961), no. 3° série, 273–304.

[8] P. D. T. A. Elliott, *Probabilistic Number Theory I.*, Springer-Verlag, New York, 1979.

[9] ———, *Probabilistic Number Theory II.*, Springer-Verlag, New York, 1980.

[10] P. Erdős and M. Kac, *On the Gaussian law of errors in the theory of additive functions*, Proc. Nat. Acad. Sci. U.S.A. **25** (1939), 206–207.

[11] ———, *The Gaussian law of errors in the theory of additive number-theoretic functions*, Amer. Journ. Math. **62** (1940), 738–742.

[12] P. Erdős and A. Wintner, *Additive arithmetical functions and statistical independence*, Amer.J.Math. **61** (1939), 713–721.

[13] J. Pintz G. Harman and D. Wolke, *A note on the Möbius and the Liouville function*, Studia Sci. Math. Hungar. **20** (1985), 295–299.

[14] J. Galambos, *Distribution of arithmetical functions. A survey.*, Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques **6** (1970), no. 4, 281–305.

[15] B. V. Gnedenko and A. N. Kolmogorov, *Limit distributions for sums of independent random variables*, Adison-Wesley Publishing Company, 1968.

[16] H. Halberstam and H.-E. Richert, *Sieve methods*, Acad. Press, London, 1974.

[17] G. Halász, *Über die Mittelwerte multiplikativer zahlentheoretischer Funktionen*, Acta Math. Acad. Sci. Hung. **19** (1968), 365–403.

[18] G.H. Hardy and S. Ramanujan, *The normal number of prime factors of a number n*, Quart. J. Pure Appl. Math. **48** (1917), 76–92.

[19] A. Hildebrand, *An Erdős-Wintner theorem for differences of additive functions*, Trans. Amer. Math. Soc. **310** (1988), no. 1, 257–276.

[20] ———, *Multiplicative functions on consecutive integers II*, Math. Proc. Camb. Phil. Soc. **103** (1988), 389–398.

[21] ———, *Additive and multiplicative functions on shifted primes*, Proc. London Math. Soc. **53** (1989), 209–232.

[22] K-H. Indlekofer, *Limiting distribution of additive functions*, Litovskii Matematicheskii Sbornik **16** (1976), no. 2, 81–91.

[23] ———, *Identities in the convolution arithmetic of number theoretical functions*, Annales Univ. Sci. Budapest., Sect. Comp. **28** (2008), 303–325.

[24] M. Kac, *Statistical independence in probability, analysis and number theory*, Wiley, New York, 1959.

[25] J. Kubilius, *Probabilistic methods in the theory of numbers*, Amer. Math. Soc. Trans. **19** (1962), 47–85.

[26] ———, *Probabilistic Methods in the Theory of Numbers*, American Mathematical Society, Providence, Rhode Island, 1964.

[27] A. V. Kumchev and D. I. Tolev, *An invitation to additive prime number theory*, Serdica. Math. J. **31** (2005), 1–74.

[28] I. Kátai, *On the distribution of arithmetical functions on the set of primes plus one*, Composito Math. **19** (1968), 278–289.

[29] ———, *On the distribution of arithmetical functions*, Acta Math. Acad. Sci. Hung. **20** (1969), no. 1-2, 69–87.

[30] I. Kátai and M. V. Subbarao, *The characterization of $n^{i\tau}$ as a multiplicative function*, Acta Math. Hungar. **81** (1998), 349–353.

[31] ———, *On the multiplicative function $n^{i\tau}$*, Studia. Sci. Math. Hungar. **34** (1998), 211–218.

[32] F. Mertens, *Ein Beitrag zur analythischen Zahlentheorie*, Crelle's Journal **78** (1874), 46–62.

[33] B. M. Phong, *On completely multiplicative functions whose values are roots of unity*, Acta Math. Hungar. **113** (2006), no. 1-2, 63–71.

[34] K. Prachar, *Primzahlverteilung*, Springer-Verlag, Berlin, 1957.

[35] Yu. V. Prokhorov, *Independence in probability theory*, `http://eom.springer.de/I/i050580.htm#i050580_00c3`, Springer online reference works.

[36] L. G. Sathe, *On a problem of of Hardy and Ramanujan on the distribution of integers having a given number of ptime factors*, J. Indian Math. Soc. **17** (1953), 63–141.

[37] A. Selberg, *Note on a paper by L. G. Sathe*, J. Indian Math. Soc. **18** (1954), 83–87.

[38] J. Siaulys and G. Stepanauskas, *On the mean value of the product of multiplicative functions with shifted arguments*, Monatsh. Math **150** (2007), 343–351.

[39] T. Tao, *Open question: The parity problem in sieve theory*, `http://terrytao.wordpress.com/2007/06/05/open-question-the-parity-problem-in-sieve-theory/`, 2007.

[40] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Cambridge University Press, 1995.

[41] N. M. Timofeev, *Multiplicative functions on the set of shifted prime numbers*, Math. USSR Izvestiya **39** (1992), no. 3, 1189–1207.

[42] N. M. Timofeev and M. B. Khripunova, *Distribution of numbers with a given number of prime divisors in progressions*, Math. Notes **55** (1994), no. 2, 204–212.

[43] ______, *The Titchmarsh problem with integers having a given number of prime factors*, Math. Notes **59** (1996), no. 4, 421–434.

[44] P. Turán, *Az egész számok prímosztóinak számáról*, Mat. Lapok **41** (1934), 103–130.

[45] ______, *On a theorem of Hardy and Ramanujan*, Journ. London Math. Soc. **9** (1934), 274–276.

[46] ______, *Über einige veralgemeinerungen eines Satzes von Hardy and Ramanujan*, Journ. London Math. Soc. **11** (1936), 125–133.

[47] E. Wirsing, *Das asymptotische Verhalten von Summen über multiplicative Funktionen*, Math. Annalen **143** (1961), 75–102.

[48] ______, *Das asymptotische Verhalten von Summen über multiplikative Funktionen II*, Acta Math. Acad. Sci Hung. **18** (1967), 411–467.

[49] D. Wolke, *Über die mittlere Verteilung der Werte zahlentheoretischer Funktionen auf Restklassen I.*, Math. Ann. **202** (1973), 1–25.

[50] D. Wolke and T. Zhan, *On the distribution of integers with a fixed number of prime factors*, Math. Z. **213** (1993), 133–147.

Publications of the author

[1] Germán, L. and Kovács, A., On number system constructions, Acta Math. Hungar. 115 (2007), no. 1-2, 155–167;

[2] Germán, L. and Kátai, I., Distribution of q-additive functions on the set of integers having k prime factors, Annales Univ. Sci. Budapest. Sect. Comp., 27 (2007), 65-74.

[3] Germán, L., The distribution of an additive arithmetical function on the set of shifted integers having k distinct prime factors, Annales Univ. Sci. Budapest. Sect. Comp., 27 (2007), 187-215.

Conference talks

[1] Germán, L., The distribution of an additive arithmetical function on the set of shifted integers having k distinct prime factors, The IVth international conference on analytic and probabilistic methods in number theory, 9/2006, Palanga

[2] Germán, L., On multiplicative functions on consecutive integers, Conference talk Numbers, Functions, Equations '08, 06/2008, Noszvaj

Summary

An arithmetical function $f : \mathbb{N} \to \mathbb{C}$ is said to be additive if $f(n \cdot m) = f(n) + f(m)$ whenever $(m, n) = 1$. Such a function is completely determined by its values on prime powers. Real valued additive function plays a special role in number theory. Let f be a real additive function, and $1 < x$. The frequency of f is defined by

$$F_x(z) := \frac{1}{x} \sum_{\substack{n \le x \\ f(n) \le z}} 1,$$

for all real z. Then $F_x(z)$ is a distribution function and we say that f has a limit distribution, if

$$F_x(z) \to F(z) \quad (x \to \infty),$$

in all z continuity points of F, where F is a distribution function as well. This type of convergence has been called weak convergence and we will denote it by

$$F_x \Rightarrow F \quad (x \to \infty).$$

It turned out, that the above property of f is equivalent to the following condition, which we call the Erdős-Wintner condition, i.e. the three series over prime numbers

$$\sum_{|f(p)|>1} \frac{1}{p}, \qquad \sum_{|f(p)| \le 1} \frac{f(p)}{p}, \qquad \sum_{|f(p)| \le 1} \frac{f^2(p)}{p}$$

converge.

This problem can be investigated in a context more general than above. For this purposes let f be an additive function and A_x be a subset of $\mathbb{N}$ such that $A_x \cap [1..x]$ is nonempty, with $1 < x$. Then the frequency of f on A_x is now defined by

$$\nu_x(n \in A_x;\ f(n) \le z) := \frac{1}{|A_x \cap [1..x]|} \sum_{\substack{n \le x \\ n \in A_x \\ f(n) \le z}} 1.$$

Then one can ask if the frequencies possess a limit law or not, that is if

$$\nu_x(n \in A_x;\ f(n) \le z) \Rightarrow F(z) \quad (x \to \infty)$$

or not with a distribution function $F(z)$. Erdős and Wintner considered this question taking A_x to be $\mathbb{N}$. Kátai and Hildebrand investigated the case $A_x = \mathcal{P}+1$, where $\mathcal{P}$ denotes the set of primes.

The main topic of the thesis is to investigate this type of distribution problems if we take

$$A_x = \{n \le x :\ \omega(n-1) = k_x\},$$

where $\omega(n)$ denotes the number of distinct prime factors of n, and $1 \le k_x \le \varepsilon(x)\sqrt{\log\log x}$ with $\varepsilon(x) \to 0 \quad (x \to \infty)$. Note that the problem of Kátai and Hildebrand corresponds to the case $k_x = 1$ (apart from the higher prime powers, which have zero relative density in the set of prime powers).

The methods used here allows us to give mean value estimations for multiplicative functions of modulus one. As for example we can prove that if there exists a real τ and a primitive character $\chi \pmod d$ for some modulus d such that

$$\sum_p \frac{|1-\overline{\chi}(p)g(p)p^{-i\tau}|}{p}$$

converge then

$$\frac{1}{|A_x|}\sum_{n\in\mathcal{A}_x} g(n+1) = \frac{x^{i\tau}}{1+i\tau}\frac{\mu(d)}{\varphi(d)}\prod_{\substack{p\le x\\ p\nmid d}}\left(1-\frac{1}{p-1}+\sum_{\alpha\ge1}\frac{f(p^\alpha)p^{-i\alpha\tau}\chi(p^\alpha)}{p^\alpha}\right) + o(1) \qquad (x\to\infty)$$

uniformly for all k with $2 \le k \le \epsilon(x)\sqrt{\log\log x}$, where $\epsilon(x) \to 0$ as $x \to \infty$.

As an application of the results discussed in the first part of this work we give an Erdős-Kac type Theorem which seems to be stated in the literature only for fixed k, and with further strong restrictions on the set A_x. As a direct consequence we obtain that

$$\nu_x\left(n \le x, \omega(n) = k : \frac{\omega(n+1)-\log\log x}{\sqrt{\log\log x}} \le z\right) \Rightarrow \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-w^2/2}dw \qquad (x\to\infty)$$

uniformly for all k with $2 \le k \le \epsilon(x)\sqrt{\log\log x}$.

Összefoglaló

Egy $f : \mathbb{N} \to \mathbb{C}$ számelméleti függvényt additívnak nevezünk, ha $f(n \cdot m) = f(n) + f(m)$ minden $(m, n) = 1$-re teljesül. Az ilyen függvényeket egyértelműen meghatározzák a prímhatvány helyeken felvett értékeik. A valós értékű additív függvények különleges szerepet töltenek be a számelméletben. Legyen f egy valós értékű additív függvény, és $0 < x$. f gyakoriságát minden valós z esetén az

$$F_x(z) := \frac{1}{x} \sum_{\substack{n \le x \\ f(n) \le z}} 1$$

utasítással értelmezük. Ekkor $F_x(z)$ egy eloszlásfüggvény, és azt mondjuk hogy f nek van határeloszlása, ha alkalmas F eloszlásfüggvény esetén

$$F_x(z) \to F(z) \quad (x \to \infty),$$

az F minden z folytonossági pontjában. Ezt a konvergenciát gyenge konvergenciának hívjuk, jelölésben:

$$F_x \Rightarrow F \quad (x \to \infty).$$

Kiderült, hogy f ezen tulajdonsága egyenértékű az ún. Erdős-Wintner feltétellel, azaz a három sor

$$\sum_{|f(p)|>1} \frac{1}{p}, \quad \sum_{|f(p)| \le 1} \frac{f(p)}{p}, \quad \sum_{|f(p)| \le 1} \frac{f^2(p)}{p}$$

konvergenciájával.

Ezt a problémát sokkal általánosabban is megfogalmazhatjuk. Legyen A_x az $\mathbb{N}$ egy olyan részhalmaza, hogy $A_x \cap [1..x]$ nem üres $1 < x$ esetén. f gyakorisága A_x-en most az

$$\nu_x(n \in A_x;\ f(n) \le z) := \frac{1}{|A_x \cap [1..x]|} \sum_{\substack{n \le x \\ n \in A_x \\ f(n) \le z}} 1$$

utasítással értelmezett. Felmerülhet a kérdés, hogy f-nek van-e határeloszlása ezen a halmazon, azaz

$$\nu_x(n \in A_x;\ f(n) \le z) \Rightarrow F(z) \quad (x \to \infty)$$

teljesül-e alkalmas $F(z)$ re. Erdős és Wintner azt a kérdést vizsgálták amikor A_x-et $\mathbb{N}$-nek vesszük. Kátai és Hildebrand az $A_x = \mathcal{P}+1$ esettel foglalkoztak, ahol $\mathcal{P}$ a prímek halmazát jelöli.

Ezen dolgozat célja hasonló eloszlásproblémák vizsgálata

$$A_x = \{n \leq x :\ \omega(n-1) = k_x\},$$

esetben ahol $\omega(n)$ az n különböző prímfaktorainak számát jelöli, és $1 \leq k_x \leq \varepsilon(x)\sqrt{\log\log x}$ ahol $\varepsilon(x) \to 0 \quad (x \to \infty)$. Észrevehetjük, hogy Kátai és Hildebrand problémája a $k_x = 1$ esetnek felel meg (a magasabb prímhatványoktól eltekintve, amelyeknek nulla a relatív sűrűsége a prímhatványok között).

Az itt használt módszerek lehetővé teszik, hogy egy abszolútértékű multiplikatív függvények középértékeiről általános érvényű következtetéseket vonjunk le. Belátjuk például, hogy ha létezik egy valós τ és egy $\chi \pmod d$ primitív karakter valamely d modulusra úgy, hogy

$$\sum_p \frac{|1-\overline{\chi}(p)g(p)p^{-i\tau}|}{p}$$

konvergál, akkor

$$\frac{1}{|A_x|}\sum_{n\in\mathcal{A}_x} g(n+1) = \frac{x^{i\tau}}{1+i\tau}\frac{\mu(d)}{\varphi(d)}\prod_{\substack{p\leq x\\ p\nmid d}}\left(1-\frac{1}{p-1}+\sum_{\alpha\geq 1}\frac{f(p^\alpha)p^{-i\alpha\tau}\chi(p^\alpha)}{p^\alpha}\right) + o(1) \qquad (x\to\infty)$$

egyenletesen minden k egészre $2 \leq k \leq \epsilon(x)\sqrt{\log\log x}$ esetén, ahol $\epsilon(x) \to 0$ ahogy $x \to \infty$.

A dolgozat első részében tárgyalt módszer alkalmazásával Erdős-Kac típusú tételek kerülnek kidolgozásra, amelyekről úgy tűnik, eddig csak rögzített k és A_x-re való megszorítás mellett szerepelnek az irodalomban. Következményként kapjuk, hogy

$$\nu_x\left(n \leq x, \omega(n) = k : \frac{\omega(n+1)-\log\log x}{\sqrt{\log\log x}} \leq z\right) \Rightarrow \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-w^2/2}dw \qquad (x\to\infty)$$

egyenletesen minden k egészre $1 \leq k \leq \epsilon(x)\sqrt{\log\log x}$ esetén.